

# Idiosyncratic Risk and the Cross-Section of Stock Returns

René Garcia, Daniel Mantilla-García and Lionel Martellini \*

May 17, 2010

## Abstract

The recent financial literature has paid considerable attention to idiosyncratic volatility. Whether average idiosyncratic volatility is a good predictor of aggregate market returns and whether idiosyncratic volatility has a positive relationship with expected returns in the cross-section are matters of active debate. We propose to measure such idiosyncratic variance by the cross-sectional variance of stock returns. Two key advantages of this measure are its observability at any frequency and its model-free nature. Previous studies have measured idiosyncratic risk only at a monthly frequency based on asset pricing models. We show that this cross-sectional measure provides a very good proxy for average idiosyncratic risk as implied by standard asset pricing models and that it predicts well aggregate returns, especially at the daily frequency. Another advantage of such a cross-sectional approach is that it can be extended to higher moments. We find that the introduction of a robust proxy for cross-sectional skewness induces a very substantial increase in the explanatory power of idiosyncratic risk for the market return.

---

\*We thank Vijay Vaidyanathan, Abraham Lioui, Stéphane Gregoir, Volker Ziemann and seminar participants at Edhec Business School for useful discussions and comments. The second author gratefully acknowledges financial support from Edhec Risk Institute. René Garcia is Professor of Finance at Edhec Business School and a research fellow at CIREQ and CIRANO. Lionel Martellini is Professor of Finance at Edhec Business School and Scientific Director of Edhec Risk Institute. Daniel Mantilla-García is PhD in Finance Candidate at Edhec Business School and Research Assistant at Edhec Risk Institute. The contact author is René Garcia, **email:** rene.garcia@edhec.edu. The address and phone number for all three authors are: **Address:** Edhec Business School, 393, Promenade des Anglais BP3116 06202 Nice cedex 3 - France. **Phone:** + 33 (0)4 93 18 99 66.

# 1 Introduction

The recent financial literature has paid considerable attention to idiosyncratic volatility. Campbell et al. (2001) and Malkiel and Xu (2002) document that idiosyncratic volatility increased over time, while Brandt et al. (2009) show that this trend completely reversed itself by 2007, falling below pre-1990s levels and suggest that the increase in idiosyncratic volatility through the 1990s was not a time trend but rather an “episodic phenomenon”. Bekaert et al. (2008) confirm that there is no trend both for the United States and other developed countries. A second fact about idiosyncratic volatility is also a source of contention. Goyal and Santa-Clara (2003) put forward that idiosyncratic volatility has forecasting power for future excess returns, while Bali et al. (2005) and Wei and Zhang (2005) find that the positive relationship is not robust to the sample chosen. Finally, while standard asset pricing models will not support the notion that idiosyncratic volatility represents a source of risk that commands a premium in equilibrium, Ang et al. (2006) find a significant relation between the returns on a cross-section of portfolios ranked according to their level of idiosyncratic volatility and their risk exposure to such a factor.

An underlying issue in all these studies is the measurement of idiosyncratic volatility. Campbell et al. (2001) use a value-weighted sum of individual firm idiosyncratic variances, computed as the variances of residuals of differences between individual firm returns and the return of an industry portfolio to which the firm belongs<sup>1</sup>. In addition to this measure, Bekaert et al. (2008) use also the individual firm residuals of a standard Fama and French three-factor model to compute a value-weighted aggregate idiosyncratic volatility<sup>2</sup>.

We revisit the issues regarding the dynamics and forecasting power of idiosyncratic variance by using instead the cross-sectional dispersion of stock returns. Through central limit arguments, we provide the conditions under which the cross-sectional variance (CSV) of stock returns converges towards the average idiosyncratic variance.<sup>3</sup> This is the case whether we use a market-cap weighted setting or an equally-weighted scheme. The advantage of this measure is obviously its observability at any frequency, while the previous approaches have used monthly measures based on time series of daily returns. A second important feature is that this measure is model-free, since we do not need to obtain residuals from a particular model to compute it.

We verify empirically that the CSV measure leads to the same conclusions that other studies (in particular Goyal and Santa-Clara (2003) and Bali et al. (2005)) have reported at

---

<sup>1</sup>This amounts to imposing unit beta restrictions in an industry-market model.

<sup>2</sup>This is also the approach followed in Ang et al. (2006).

<sup>3</sup>Goyal and Santa-Clara (2003) argue informally that their measure can be interpreted as a measure of cross-sectional dispersion of stock returns, but do not establish a formal link between the two. In the practitioners’ literature (see DiBartolomeo (2006)), cross-sectional dispersion of returns is called variety of returns and is used in estimating correlations of assets, risk management and performance analysis.

the monthly frequency. We report new results at the daily frequency. Specifically, we show that the predictive power of idiosyncratic volatility is much stronger both quantitatively and statistically at the daily frequency than at the monthly frequency. We also estimate a regime switching model for CSV time series at both the daily and monthly frequencies and find remarkably coherent results in terms of parameter estimates. This is not the case when we use the idiosyncratic variance measure based on the Fama-French residuals as in Ang et al. (2006) and Bekaert et al. (2008).

Another advantage of the cross-sectional approach is that it extends naturally to higher moments. Ang et al. (2006) rationalize the relation between idiosyncratic volatility and average returns by the fact that firms with greater sensitivities to aggregate volatility should have larger idiosyncratic volatilities relative to the Fama-French model since they will show up in the residuals of the latter. By the same token, since skewness has been shown to be a priced factor by Harvey and Siddique (2000), a measure of cross-sectional skewness of returns may have some forecasting power for average future returns.<sup>4</sup> We provide conditions under which cross-sectional skewness is a good measure of aggregate idiosyncratic skewness and use a robust measure of cross-sectional skewness in a predictive regression together with CSV to show that skewness is a strong predictor of future returns. The adjusted  $R^2$  of the regression climbs to 6% from a 0.6% with only CSV at daily horizon and to 4.6% from 0.8% at monthly horizon.

Our results indicate that the relation is much stronger and stable across periods between the equally-weighted measure of aggregate idiosyncratic volatility and the returns on the equally-weighted index than for the market-cap weighted equivalents. Economic sources of heterogeneity between firms, as diverse as they can be, are better reflected in an equally weighted measure, all other things being equal. This argument is consistent with previous findings in Bali et al. (2005), who argue that the relationship between equal-weighted average idiosyncratic risk and the market-cap weighted index on the sample ending in 1999:12 is mostly driven by small stocks traded in the NASDAQ. Of course, when the bubble burst, the market capitalization of dot.com small firms was relatively more affected causing the relationship to break down in 2000 and 2001. This effect is not prevalent in an equally-weighted index, for which the relationship remains strong.

The statistical significance of the moments of the cross-sectional distribution in these predictive regressions of future returns is not the same as the cross-sectional pricing of stocks or portfolios. However, as emphasized in Ang et al. (2006) and Goyal and Santa-Clara (2003), the two pieces of evidence are related. Theoretical rationalizations of a positive relation between idiosyncratic volatility and expected returns can be found in the asset

---

<sup>4</sup>Boyer et al. (2009) find a significant relationship between expected idiosyncratic skewness and future returns at the stock level.

pricing literature. The Levy (1978), Merton (1987), Markowitz (1990) and Malkiel and Xu (2002) extensions of the CAPM argue that an important portion of investors' portfolios may differ from the market. Their holdings may be affected by corporate compensation policies, borrowing constraints, heterogeneous beliefs and include *non-traded* assets that add background risk to their traded portfolio decisions (e.g. human capital and private businesses). Levy (1978), Merton (1987) and Malkiel and Xu (2002) pricing models relate stock returns to their beta with the market and their beta to market-wide measures of idiosyncratic risk. Furthermore, Campbell et al. (2001) mentioned that some investors try to approximate a well-diversified portfolio using the rule of thumb of holding between 20 to 30 stocks hoping to eliminate all idiosyncratic risk, which may depend on the current level of average idiosyncratic risk. Under these considerations, investors can be affected by changes in idiosyncratic volatility just as much as by changes in market volatility. More recently, Guo and Savickas (2008) argue that changes in average idiosyncratic volatility provide a proxy for changes in the investment opportunity set and that this proxy is closely related to the book-to-market factor. Ang et al. (2006) and Ang et al. (2008) find results that are opposite to these theories since stocks with high idiosyncratic volatility have low average returns but cannot fully rationalize this result. However, Huang et al. (2009) find that the negative sign in the relationship between idiosyncratic variance and expected returns at the stock level becomes positive after controlling for return reversals. Similarly, Fu (2009) documents that high idiosyncratic volatilities of individual stocks are contemporaneous with high returns, which tend to reverse in the following month.

Goyal and Santa-Clara (2003) also point out that considering individual stocks as a proxy for idiosyncratic income of investors, allows us to interpret average stock risk as a measure of cross-sectional variance of income shocks among investors. This would provide supporting evidence for models based on heterogeneity to explain the market risk premium, such as Constantinides and Duffie (1996), where income shocks must be persistent and their average variance counter-cyclical (which is the case for the average stock variance). Alternative explanations of the relation between idiosyncratic risk and return are the firm's assets' call-option interpretation by Merton (1974) where equity is a function of total volatility as in Black and Scholes (1973) as well as Barberis et al. (2001) prospect theory asset pricing model with loss aversion over (owned) individual stock's variance.

The frequency at which predictive regressions are run has an impact on the results, since at lower frequencies the sign of the relation between idiosyncratic volatility and future returns changes. Guo and Savickas (2008) find that at a quarterly frequency the relation is negative, in contrast with the positive monthly and daily results that we document. They attribute this puzzling result to the fact that the value-weighted volatility is negatively

correlated with the consumption-wealth ratio.<sup>5</sup>

The rest of the paper is organized as follows. In Section 2, we provide a formal argument for choosing the cross-sectional variance of returns as a measure of average idiosyncratic volatility, explore its properties and the assumptions behind its use, and compare it to other measures formerly selected in the literature. Section 3 provides an empirical implementation of the concept by studying its time-series behavior and the presence of regimes and counter-cyclical property again in comparison with other measures. In section 4, we naturally extend the cross-sectional concept and formally explore the link between cross-sectional and idiosyncratic skewness. In Section 5, we provide new results on the predictability of returns by idiosyncratic volatility and skewness. Section 6 sketches an interpretation of our results in relation with previous evidence. Section 7 concludes and a technical appendix collects proofs and more formal derivations.

## 2 Idiosyncratic Risk and the Cross-Sectional Variance (CSV) of Realized Returns

In this section, we provide the conditions and assumptions under which the cross-sectional variance of stock returns provides a good measure of idiosyncratic volatility. We discuss some of its properties and advantages with respect to competing measures in the literature. Previous studies that have used the CSV to capture idiosyncratic volatility, have not provided a thorough discussion about it can be considered as a “good” measure of (average) idiosyncratic variance and the implications of the different assumptions that stand behind its use.

### 2.1 A Formal Argument for Choosing the CSV as a Measure of Idiosyncratic Volatility

To set a framework for measuring idiosyncratic risk, we assume without loss of generality the following conditional single factor model for excess stock returns of an asset or portfolio  $i$ :

$$r_{it} = \beta_{it}F_t + \varepsilon_{it}. \quad (1)$$

Given  $T$  observations of the returns and the factor, one can use the residuals of a regression to obtain a measure of the idiosyncratic variance of asset  $i$  by:  $\sigma_i^2 = \frac{1}{T} \sum_{t=1}^T \varepsilon_{it}^2$ . An aggregate measure of idiosyncratic variance over the  $T$  observations (say a month) can be obtained by averaging over  $N$  assets these individual idiosyncratic variances. This is the

---

<sup>5</sup>This tends to be consistent with the findings of Ang et al. (2006) but Fu (2009) argues that these findings are largely driven by the return reversal of stocks that have high idiosyncratic volatilities.

approach that has been followed by most papers with observations of the returns at a daily frequency to compute monthly idiosyncratic variances. We propose instead to measure at each time  $t$  the cross-sectional variance of observed asset or portfolio returns. We motivate below under simplifying assumptions that such a cross-sectional measure provides a very good approximation to aggregate idiosyncratic variance.

Let  $N_t$  be the total number of stocks in the universe at day  $t$ ,  $(w_t)_{t \geq 0}$  a weight vector process. The return on the portfolio defined by the weight vector process  $(w_t)$  is denoted by  $r_t^{(w_t)}$  and given by:

$$r_t^{(w_t)} = \sum_{i=1}^{N_t} w_{it} r_{it}. \quad (2)$$

In general, the factor  $F_t$  can be different from the market index  $r_t^{(w_t)}$ . We restrict our attention to non-trivial weighting schemes, ruling out situations such that the index is composed by a single stock. We also restrict the weights to be positive at every given point in time. Hence, a weighting scheme  $(w_t)$ , is a vector process which satisfies  $0 < w_{it} < 1 \forall i, t$ . This condition seems reasonable since our focus is to measure average idiosyncratic risk in the market.

The cross-sectional variance measure is defined as follows:

**Definition (*CSV*):** The ***cross-sectional variance*** measure under the weighting scheme  $(w_t)$ , denoted by  $CSV_t^{(w_t)}$ , is given by

$$CSV_t^{(w_t)} = \sum_{i=1}^{N_t} w_{it} \left( r_{it} - r_t^{(w_t)} \right)^2. \quad (3)$$

A particular case of interest is the ***equally-weighted CSV***, denoted by  $CSV_t^{EW}$  and determined by the weighting scheme  $w_{it} = 1/N_t \forall i, t$ :

$$CSV_t^{EW} = \frac{1}{N_t} \sum_{i=1}^{N_t} \left( r_{it} - r_t^{EW} \right)^2, \quad (4)$$

where  $r_t^{EW}$  is the return on the equally-weighted portfolio.

If we denote by  $c_{it}$  be the market capitalization of stock  $i$  at the beginning of the month corresponding to day  $t$ ,  $C_t = \sum_{i=1}^{N_t} c_{it}$  the total market capitalization and  $r_t^{CW}$  the return on the market capitalization-weighted portfolio, let the ***cap-weighted CSV*** be define as:

$$CSV_t^{CW} = \sum_{i=1}^{N_t} w_{it}^{CW} \left( r_{it} - r_t^{CW} \right)^2 \quad (5)$$

where  $w_{it}^{CW} = \sum_{i=1}^{N_t} \frac{c_{it}}{C_t}$ .

The advantage of such a cross-sectional measure is obviously its observability at any frequency, while the previous approaches have used monthly measures based on time series of daily returns. A second important feature is that this measure is model-free, since we do not need to obtain residuals from a particular model to compute it.

While Goyal and Santa-Clara (2003) and Wei and Zhang (2005) consider the equal-weighted CSV in conjunction with other measures, they do not provide conditions under which it can be interpreted as a proxy for idiosyncratic variance. To study the nature of CSV, we will make three assumptions and use them at various stages to derive the properties of the CSV as an estimator of idiosyncratic variance.

**Assumption 1**  $\forall i, t, \varepsilon_{it}$  is independently distributed in time with homogeneous second moment across assets, with

$$\begin{aligned} E(\varepsilon_{it}) &= 0. \\ E(\varepsilon_{it}^2) &= \sigma_\varepsilon^2(t). \\ \text{Corr}(F_t, \varepsilon_{it}) &= 0. \end{aligned}$$

This assumption is consistent with the usual assumptions made in factor models. The next assumption, made by previous authors in the idiosyncratic risk literature (see Campbell et al. (2001) and Goyal and Santa-Clara (2003) in particular), reduces to zero the cross-sectional dispersion of betas.

**Assumption 2**  $\beta_{it} = \beta_t = 1 \forall i, t$

Later in this section we discuss the implications in terms of bias of this assumed homogeneity in betas and in the next section we will assess the importance of this bias empirically.

For the next proposition, we consider the simple case where the factor  $F_t$  is the market portfolio defined by the set of weights,  $w_t$ , hence  $F_t = r_t^{(w_t)}$ . Under assumptions 1 and 2, we the following Corollary draws a direct relationship between the dynamics of the cross-sectional variance of realized returns and the dynamics of idiosyncratic variance.

**Proposition 1 (CSV as proxy for idiosyncratic variance):** *with a positive weighting scheme and under assumptions 1 and 2, we have that,*

$$CSV_t^{(w_t)} = \sum_{i=1}^{N_t} w_{it} \left( r_{it} - r_t^{(w_t)} \right)^2 \xrightarrow{N_t \rightarrow \infty} \sigma_\varepsilon^2(t) \text{ almost surely} \quad (6)$$

*(E[CSV] as average idiosyncratic variance measure): Moreover,*

$$E \left[ CSV_t^{(w_t)} \right] = \sum_{i=1}^{N_t} w_{it} \sigma_{\varepsilon_i}^2(t) \quad (7)$$

This result holds even for the more general case where the factor(s) is not necessarily the market portfolio defined by the weighting scheme  $w_t$ . In Appendix A we present the proof of Proposition 1 in this more general case. Appendix A.1 presents a similar derivation to that proof for the equally-weighted  $CSV$ , which follows in a similar but simpler manner. For the simple case above, the proof is quite simple.

**Proof** From the single factor decomposition (1), it follows that the difference between any individual return and the return on the market factor is

$$r_{it} - r_t^{(w_t)} = (\beta_{it} - 1)r_t^{(w_t)} + \varepsilon_{it}.$$

The homogeneous beta assumption implies:

$$r_{it} - r_t^{(w_t)} = \varepsilon_{it}. \quad (8)$$

Replacing expression (8) in the definition of the CSV (3) we have

$$CSV_t^{(w_t)} = \sum_{i=1}^{N_t} w_{it} \left( r_{it} - r_t^{(w_t)} \right)^2 = \sum_{i=1}^{N_t} w_{it} \varepsilon_{it}^2. \quad (9)$$

For independently distributed  $\varepsilon_i$  with homogeneous idiosyncratic variance across-stocks, the approximation  $\sum_{i=1}^{N_t} w_{it} \varepsilon_i^2(t) \simeq \sigma_\varepsilon^2(w_t)$ , follows from the definition  $E[\varepsilon_i^2] \equiv \sigma_\varepsilon^2$  of Assumption 1, which implies

$$CSV_t^{(w_t)} = \sum_{i=1}^{N_t} w_{it} \varepsilon_{it}^2 \simeq \sigma_\varepsilon^2(w_t) \quad \blacksquare \quad (10)$$

The proof of the second part of the proposition follows simply from taking the expectation in equation (9).

Proposition 1 is a rather intuitive argument to motivate the  $CSV$  as a measure of idiosyncratic variance in general. It is an approximation that relies on the assumption that all idiosyncratic variances are equivalent across stocks ( $E[\varepsilon_i^2] = \sigma_\varepsilon^2$ ). This approximation allows us to use *only* instantaneous information from the cross-section to estimate idiosyncratic risk in the market. Although, from a purely statistical standpoint it is more natural to give equal probabilities to every observation in the sample (which will be the case for the equal-weighted CSV), sometimes a different type of weighting scheme (or weighted average) is of interest for economic or comparison reasons, i.e. the market capitalization weighting scheme.

The second part of the proposition draws a direct relationship between the *expected value* of the CSV and the *average* idiosyncratic variance of the market, which actually *relaxes* the assumption of homogeneous idiosyncratic variances across stocks.



Notice that the weighting scheme defining the CSV and the weighted average are assumed to be the same. The proof for the more general case where the factor is not exactly the market portfolio defined by the weighting scheme  $w_t$ , uses the additional assumption of a *strict* factor model and an arbitrarily large number of stocks, as described in detail in section 2.3.

**Assumption 3** *A strict factor model implies*

$$\text{Corr}(\varepsilon_{it}, \varepsilon_{jt}) \equiv \rho_{ijt}^\varepsilon = 0 \quad \forall i \neq j, t$$

Assumption 3 is similar to the single index, or diagonal, model of Sharpe (1963) and is also used by statistical factor models. It implies that all commonalities are explained by the factor model in place. One should notice that the very definition of idiosyncratic risk relies precisely on this assumption about the factor model. Assuming that the model is the “true” one implies that the “true” idiosyncratic risk is the one measured with respect to that model, which in turn implies that no commonalities should be left and neither “matter” since no residual correlations should be significant. Relaxing the assumption that the factor model in place is strict means that idiosyncratic risk is not *truly* “idiosyncratic”, and hence it should rather be interpreted as a relative measure with respect to the benchmark model. In section 2.3, where we study the properties of the CSV, we illustrate in a more concrete way the implications for the CSV when this assumption does not hold.

In the next section, we explore the bias associated with the assumption of homogeneous betas.

## 2.2 Implications of the Homogeneous Beta Assumption

The assumption that  $\beta_{it} = \beta_t$  for all  $i$  is obviously a simplistic one. We now look at the CSV when this assumption is not introduced. Noting that  $\sum_{i=1}^{N_t} w_{it} \beta_{it} = 1$ , we have

$$CSV_t^{(w_t)} = \sum_{i=1}^{N_t} w_{it} (r_{it} - F_t)^2 \quad (11)$$

$$= \sum_{i=1}^{N_t} w_{it} [(\beta_{it} - 1) F_t + \varepsilon_{it}]^2 \quad (12)$$

$$= F_t^2 \sum_{i=1}^{N_t} w_{it} (\beta_{it} - 1)^2 + \sum_{i=1}^{N_t} w_{it} \varepsilon_{it}^2 + 2F_t \sum_{i=1}^{N_t} w_{it} (\beta_{it} - 1) \varepsilon_{it} \quad (13)$$

We now look at the expectation of the CSV. Define  $CSV_t^\beta = \sum_{i=1}^{N_t} w_{it} (\beta_{it} - 1)^2$  as the cross-sectional variance of stock betas. Applying the expectation operator in equation (13)

and assuming  $F_t$  orthogonal to  $\beta_t$  we get,

$$E \left[ CSV_t^{(w_t)} \right] = E \left[ F_t^2 CSV_t^\beta \right] + \sum_{i=1}^{N_t} w_{it} E \left[ \varepsilon_{it}^2 \right] + 2E \left[ F_t \right] \sum_{i=1}^{N_t} w_{it} (\beta_{it} - 1) E \left[ \varepsilon_{it} \right]$$

Assuming  $E[\varepsilon_{it}] = 0$  yields,

$$E \left[ CSV_t^{(w_t)} \right] = E \left[ F_t^2 CSV_t^\beta \right] + \sum_{i=1}^{N_t} w_{it} \sigma_{\varepsilon_{it}}^2 \quad (14)$$

The first term in equation (14) represents a (positive) bias for the CSV as an estimator of average idiosyncratic variance, introduced by the betas' cross-sectional dispersion. A similar derivation for the bias in the more general case where the factor is not the same as the market portfolio defined by  $w_t$  is provided in Appendix B. Under the additional assumption of a *strict factor model*, the bias of the CSV turns out to be the same expression. The implication of not assuming a strict factor model are shown in section 2.3.

In section 3.1 we measure the bias of the CSV as a measure of average idiosyncratic risk with respect to standard asset pricing models in the literature (CAPM and Fama-French) and find that the size of the CSV's bias is negligible. As we will see, although the cross-sectional dispersion of betas has a non-negligible magnitude, once it is multiplied by the squared of the return of the market portfolio its relative size with respect to the level of idiosyncratic risk becomes very small.

## 2.3 CSV properties

In this section, we study the properties of the CSV estimator. These properties are of particular interest to compare the equal-weighted and cap-weighted versions of the *CSV*. The equal-weighted *CSV* appears as a *consistent* and *asymptotically efficient* estimator of idiosyncratic variance in the class of CSV estimators defined under any positive weighting scheme.

### 2.3.1 Expectation and Bias

Proposition 1 follows only in the case where  $F_t = r^{(w_t)}$ . When it is not the case, the CSV estimator is a biased measure of average idiosyncratic variance, where the bias is proportional to the concentration index implied by the set of weights  $w_t$ . The bias reaches a minimum for the equally-weighted scheme and tends to zero as the number of stocks grows to infinity. This result is illustrated in Corollary 1 below.

**Lemma 1** (*CSV<sup>EW</sup> asymptotically unbiased estimator*): For an arbitrarily large number of stocks the equal-weighted CSV is an asymptotically unbiased estimator of average idiosyncratic variance. For a small number of stocks, it has the smallest bias among the class of estimators defined by the CSV with a weighting scheme  $0 < w_{it} < 1$ .

$$E [CSV_t^{EW}] \xrightarrow{N_t \rightarrow \infty} \frac{1}{N_t} \sum_{i=1}^{N_t} \sigma_{\varepsilon_{it}}^2$$

**Proof** of Lemma 1

In the more general case where the factor is not necessarily the portfolio defined by  $w_t$ , under the factor model decomposition (1) and equation (2) and using the homogeneous beta assumption, we have

$$\begin{aligned} r_{it} - r_t^{(w_t)} &= \left( \beta_{it} - \sum_{j=1}^{N_t} w_{jt} \beta_{jt} \right) F_t + \varepsilon_{it} - \sum_{j=1}^{N_t} w_{jt} \varepsilon_{jt} \\ r_{it} - r_t^{(w_t)} &= \varepsilon_{it} - \sum_{j=1}^{N_t} w_{jt} \varepsilon_{jt} \end{aligned} \quad (15)$$

Replacing result (15) in equation (3) we have

$$\begin{aligned} CSV_t^{(w_t)} &= \sum_{i=1}^{N_t} w_{it} \left( \varepsilon_{it} - \sum_{j=1}^{N_t} w_{jt} \varepsilon_{jt} \right)^2 \\ &= \sum_{i=1}^{N_t} w_{it} \varepsilon_{it}^2 + \sum_{i=1}^{N_t} w_{it} \left( \sum_{j=1}^{N_t} w_{jt} \varepsilon_{jt} \right)^2 - 2 \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} w_{jt} w_{it} \varepsilon_{it} \varepsilon_{jt} \end{aligned}$$

and noting that

$$\left( \sum_{j=1}^{N_t} w_{jt} \varepsilon_{jt} \right)^2 = \sum_{j=1}^{N_t} w_{jt}^2 \varepsilon_{jt}^2 + 2 \sum_{i \neq j}^{N_t-1} w_{it} w_{jt} \varepsilon_{it} \varepsilon_{jt} = \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} w_{jt} w_{it} \varepsilon_{it} \varepsilon_{jt}$$

we get

$$CSV_t^{(w_t)} = \sum_{i=1}^{N_t} w_{it} \varepsilon_{it}^2 - \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} w_{jt} w_{it} \varepsilon_{it} \varepsilon_{jt} \quad (16)$$

By definition,  $E[\varepsilon_{it} \varepsilon_{jt}] = \rho_{ijt}^\varepsilon \sigma_{\varepsilon_{it}} \sigma_{\varepsilon_{jt}}$  and  $E(\varepsilon_{it}^2) = \sigma_{\varepsilon_{it}}^2$ . Applying the Expectation operator

in equation (16)<sup>6</sup> we get,

$$E \left[ CSV_t^{(w_t)} \right] = \sum_{i=1}^{N_t} w_{it} \sigma_{\varepsilon_{it}}^2 - \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} w_{it} w_{jt} \rho_{ijt}^{\varepsilon} \sigma_{\varepsilon_{it}} \sigma_{\varepsilon_{jt}} \quad (17)$$

Separating variance and covariance terms in equation (17) we have,

$$E \left[ CSV_t^{(w_t)} \right] = \sum_{i=1}^{N_t} w_{it} \sigma_{\varepsilon_{it}}^2 - \sum_{i=1}^{N_t} w_{it}^2 \sigma_{\varepsilon_{it}}^2 - \sum_{i=1}^{N_t} \sum_{j \neq i}^{N_t} w_{it} w_{jt} \rho_{ijt}^{\varepsilon} \sigma_{\varepsilon_{it}} \sigma_{\varepsilon_{jt}} \quad (18)$$

For a strict factor model as in Assumption 3<sup>7</sup> it follows that

$$E \left[ CSV_t^{(w_t)} \right] = \sum_{i=1}^{N_t} w_{it} \sigma_{\varepsilon_{it}}^2 - \sum_{i=1}^{N_t} w_{it}^2 \sigma_{\varepsilon_{it}}^2 \quad (19)$$

The second term in (19) implies that the CSV would tend to underestimate average idiosyncratic variance. However, for the equal weighted CSV we can derive the following result in this context.

Considering the case where  $w_{it} = 1/N_t \forall i$  in (19) simplifies to

$$E \left[ CSV_t^{EW} \right] = \left( 1 - \frac{1}{N_t} \right) \frac{1}{N_t} \sum_{i=1}^{N_t} \sigma_{\varepsilon_{it}}^2$$

for  $N \rightarrow \infty$

$$E \left[ CSV_t^{EW} \right] \xrightarrow{N_t \rightarrow \infty} \frac{1}{N_t} \sum_{i=1}^{N_t} \sigma_{\varepsilon_{it}}^2 \quad \blacksquare$$

### 2.3.2 Variance and Efficiency

Consider equation (16) in matrix form. Let  $w_t$  and  $\varepsilon_t$  be column vector of the weighting scheme and residuals respectively and  $\Omega_t = w_t w_t'$ ,  $\Lambda_t = \text{diag}(w_t)$ ,  $N_t \times N_t$  matrices,

$$CSV_t^{(w_t)} = \varepsilon_t' \Lambda_t \varepsilon_t - \varepsilon_t' \Omega_t \varepsilon_t \quad (20)$$

---

<sup>6</sup>Equation (16) can be re-written in matrix form: Let  $\Omega_t = w_t w_t'$  be the  $N_t \times N_t$  matrix composed by a weighting scheme  $w_t$  and  $\varepsilon$  the column vector of residuals,

$$CSV_t^{(w_t)} = w_t' \varepsilon^2 - \varepsilon_t' \Omega_t \varepsilon$$

<sup>7</sup>Relaxing the strict factor model and assuming without loss of generality that  $\rho_{ijt}^{\varepsilon} = \rho_i^{\varepsilon} \forall i \neq j$  we obtain:  $E \left[ CSV_t^{(w_t)} \right] = \sigma_{\varepsilon_t}^2 \left( 1 - \sum_{i=1}^{N_t} w_{it}^2 \right) (1 - \rho_i^{\varepsilon})$ . Therefore, depending on the sign of  $\rho_i^{\varepsilon}$  the average idiosyncratic variance can be over or under estimated. Moreover, the bias is proportional to the concentration index (also known as the Herfindahl index) of the corresponding CSV, given by  $\sum_{i=1}^{N_t} w_{it}^2$ . In section 3.1 we measure and estimate the size of the CSV bias through comparison with the average idiosyncratic variance implied by CAPM and the three factor Fama-French model.

We shall also consider the simpler case where the factor is given by the portfolio defined by the set of weights,  $w_t$ , in matrix form

$$CSV_t^{(w_t)} = \varepsilon_t' \Lambda_t \varepsilon_t \quad (21)$$

We know the variance of quadratic forms for multivariate normal (distributed) variables<sup>8</sup>; denoting  $\Sigma^\varepsilon$  the variance covariance matrix of the residuals we have<sup>9</sup>,

$$Var(\varepsilon_t' \Lambda_t \varepsilon_t) = 2tr(\Lambda_t \Sigma_t^\varepsilon \Lambda_t \Sigma_t^\varepsilon) \quad (22)$$

$$Var(\varepsilon_t' \Omega_t \varepsilon_t) = 2tr(\Omega_t \Sigma_t^\varepsilon \Omega_t \Sigma_t^\varepsilon) \quad (23)$$

$$Cov(\varepsilon_t' \Lambda_t \varepsilon_t; \varepsilon_t' \Omega_t \varepsilon_t) = 2tr(\Lambda_t \Sigma_t^\varepsilon \Omega_t \Sigma_t^\varepsilon) \quad (24)$$

due to the quadratic form of the CSV, in order to analyze its variance we further assume normality of the residuals,

**Assumption 4**

$$\varepsilon \sim N(0, \Sigma^\varepsilon)$$

In the context where  $F_t = r^{(w_t)}$  and further taking Assumption 4 we are able to derive the following Corollary,

**Lemma 2** *As the number of stocks in the universe increases, the variance of the equally-weighted CSV decreases and collapses to zero in the limit*

$$Var\left(CSV_t^{(EW)}\right) \xrightarrow{N_t \rightarrow \infty} 0$$

**Proof** of Lemma 2. Replacing result (22) in equation (21) we get,

$$Var\left(CSV_t^{(w_t)}\right) = 2 \sum_{i=1}^{N_t} w_{it}^2 \sigma_{\varepsilon_{it}}^4 \quad (25)$$

consider the equal-weighted CSV for which  $w_{it} = 1/N_t \forall i$

$$Var\left(CSV_t^{(w_t)}\right) = \frac{2}{N_t} \left( \frac{1}{N_t} \sum_{i=1}^{N_t} \sigma_{\varepsilon_{it}}^4 \right)$$

when  $N_t \rightarrow \infty$  it follows

$$Var\left(CSV_t^{(EW)}\right) \xrightarrow{N_t \rightarrow \infty} 0 \quad \blacksquare \quad (26)$$

In Appendix C, we treat the more general case where the factor is not necessarily the portfolio defined by  $w_t$ .

---

<sup>8</sup>See for instance Kachman (1999)

<sup>9</sup>tr stands for the trace of a matrix, which is the sum of the diagonal terms.

**Proposition 2** (*CSV<sup>EW</sup> consistent estimator*): *The equal-weighted CSV is a consistent estimator of idiosyncratic variance.*

Proposition 2 follows from Corollaries 1 and 2 above.

We have shown that in the simple case where the factor is equal to the portfolio with the same weighting scheme defining the CSV, the equal-weighted CSV is a *consistent* estimator of average idiosyncratic variance in the market.

For the more general case where the factor is not necessarily given by the portfolio defined by the CSV weighting scheme, we find for a strict factor model that the variance and the bias of the CSV are proportional to the level of concentration implied by the corresponding vector of weights. In this context, the equal-weighted CSV is the most efficient estimator in the class of estimators given by the CSV under a weighting scheme satisfying  $0 < w_i < 1 \forall i$ .

Proposition 1 and subsequent results in this section provide motivation to use the *expected value* of the CSV as a measure of *average* idiosyncratic variance. This implies that a window of CSV observations should be used to estimate its expected value (which is typically done by taking an average over a month of daily observations). However, Proposition 1 also motivates the CSV as a proxy for idiosyncratic risk in the market, which allows us to use even instantaneous observations of the CSV for this purpose.

## 2.4 Competing Measures of Idiosyncratic Risk

Broadly speaking, our measure of idiosyncratic risk has three main advantages with respect to existing measures: it is instantaneous, readily observable, and leads to a straightforward extension to higher moments (see section 4). In what follows, we provide information regarding existing measures that have been used in the literature, and which will be used for comparison purposes in subsequent sections of the paper.

The standard approach consists of considering idiosyncratic variance either relative to the CAPM and or to the Fama-French model (Fama and French (1993)):

$$r_{it} = b_{0it} + b_{1it}MKT_t + b_{2it}SMB_t + b_{3it}HML_t + \varepsilon_{it}^{FF} \quad (27)$$

where  $r_{it}$  denotes the excess return at time  $t$  of stock  $i$ ,  $MKT$  is the excess return on the market portfolio,  $SMB$  is the size factor and  $HML$  is the value factor<sup>10</sup>.

The idiosyncratic variance for asset  $i$  is the variance of the residual of the regression, that is,  $\sigma^2(\varepsilon_{it}^{FF})$ . To obtain an estimate for average idiosyncratic variance, Bekaert et al. (2008) and Wei and Zhang (2006) use a market capitalization weighting

$$FF_t^{CW} = \sum_{i=1}^{N_t} w_{it} \sigma^2(\varepsilon_{it}^{FF}). \quad (28)$$

---

<sup>10</sup>Data on Fama-French factors and the risk-free rate are obtained from Kenneth French data library.

For comparison purposes we also look at the equally-weighted average of FF idiosyncratic variance in what follows. An alternative approach to average (*mostly*) idiosyncratic risk estimation has been suggested by GS, with a measure given by:

$$GS_t^{EW} = \frac{1}{N_t} \sum_{i=1}^{N_t} \left[ \sum_{d=1}^{D_t} r_{id}^2 + 2 \sum_{d=2}^{D_t} r_{id} r_{id-1} \right], \quad (29)$$

where  $r_{id}$  is the return on stock  $i$  in day  $d$  and  $D_t$  is the number of trading days in month  $t$ .<sup>11</sup>

Campbell et al. (2001) propose yet an alternative measure of average idiosyncratic variance, under a very particular setting which allows one to avoid running regressions each period. However, their measure is not instantaneous since a window of data is still needed to estimate individual variances. In what follows, we do not repeat the analysis with this measure because Bekaert et al. (2008) have shown that it is very closely related to the measure obtained from standard asset pricing models. In particular, Bekaert et al. (2008) find a correlation of 98% between the measure of Campbell et al. (2001) and the FF-based one. They also find that most structural breaks are identical for both measures.

### 3 Empirical Implementation

In order to perform an empirical analysis of our measure of idiosyncratic risk, we collect daily US stock returns (common equity shares only) and their market capitalization from CRSP (daily returns availability). We also extract the FF factors and the one-month Treasury bill from Kenneth French web-site data library for the same sample period. Each month, we drop stocks with missing returns and with non-positive market capitalization at the beginning of the month. Then we estimate cap-weighted idiosyncratic variance as in equation (28), and also the equal-weighted version every month<sup>12</sup>. Similarly, we estimate the cap-weighted and equal weighted average idiosyncratic variance relative to CAPM. We also estimate the GS average variance measure as in equation (29) and its cap-weighted version. Finally, we estimate on a daily basis the equal and cap-weighted (sometimes abbreviated by EW and CW) versions of the  $CSV$  as in equations (4) and (5). In order to construct the monthly series of our cross-sectional measure, we estimate the average of the daily series at the end of each month. For comparison purposes we also estimate the FF-based average idiosyncratic variance (EW and CW) on a daily basis using a rolling window sample of one month. We annualize all figures in order to compare results obtained with daily and monthly data.

---

<sup>11</sup>As in GS, when the second term makes the estimate negative, it is ignored. This measure has been originally used in French et al. (1987)

<sup>12</sup>We use previous period market capitalization and assume it is constant within the month.

Following Bekaert et al. (2008), we also fit a regime-switching model to the monthly and daily series, which allows us to further compare the different measures.

### 3.1 Measuring the CSV bias

Previous researchers analyzing average idiosyncratic volatility have assumed homogeneous betas across stocks. As we uncovered in section 2.2, non-homogeneous betas introduce a positive bias on the CSV as an estimator of average idiosyncratic variance, which is given by the first term in equation (14). We now measure the impact of this bias taking as a benchmark model the CAPM<sup>13</sup>.

First we estimate the bias every month in the sample using beta estimates for every stock and the observed market portfolio return (both equal and cap weighted versions). More importantly we look at the size of the ratio of  $E \left[ F_t^2 CSV_t^\beta \right]$  (the bias) to the average idiosyncratic variance, also measured with respect to the CAPM<sup>14</sup>.

Table 1 presents a summary of the unconditional distribution of the cross-sectional dispersion of betas, its product with the squared return of the market portfolio (hence the bias itself) and the proportion of this bias with respect to the average idiosyncratic variance at the end of every month. Although the cross-sectional dispersion of betas is sizable, once it is multiplied by the squared return of the market portfolio, the size of the bias remains fairly small: the median of the distribution of the proportion of the bias with respect to the average idiosyncratic variance,  $\frac{F_t^2 CSV_t^\beta}{\sigma_{\varepsilon_t}^2}$ , is 0.65% for the equal-weighted and 0.31% for the cap-weighted measures and the quantile 97.5 of this time series during the whole sample period (July 1963 to December 2006) is 6.67% and 2.81% correspondingly. The average of this proportion is 1.61% for the equal-weighted and 0.71% for the cap-weighted one.

On the other hand, the formal discussion about the properties of the CSV as a measure of idiosyncratic variance on section 2.3 also uncovered the fact that, when the factor model is not exactly equal to the portfolio determined by the weighting scheme under which the CSV is defined, another bias (but negative in sign) coming from the CSV's weighting scheme *concentration*<sup>15</sup> is also introduced. Equation (19) predicts two properties about this *weighting bias*: first, it should be negative and minimal for an equally-weighted scheme. Second, it should be very small for a high number of stocks. It should be noted that this bias is somehow competing with the beta's cross-sectional dispersion bias, because they have opposite signs. The beta-bias then is more likely to dominate the weighting one when using an equal-weighting scheme. Additionally, if the biases are measured with respect to

---

<sup>13</sup>Using as the benchmark multi-factor models, such as the Fama-French one, is likely to increase the size of the bias, since a similar term would appear for each additional factor.

<sup>14</sup>This is measured similar as in equation (28) for the corresponding weighted average.

<sup>15</sup>Concentration measured as the Herfindahl index



idiosyncratic risk implied by a multi-factor model, they are more likely to have a higher impact.

In order to determine the impact of these biases (both the betas and the weighting one), we measure the total bias as the intercept of a regression of the CSV on the average idiosyncratic variance estimated with respect to the CAPM or the Fama-French three-factor model, as follows

$$CSV_t^{wt} = bias + \psi \sigma_{model}^2(w_t) + \zeta_t \quad (30)$$

where  $w_t$  would be either equal-weighted or market-cap weighting schemes and  $model$  would be either the CAPM or the Fama-French three factor model.

Table 2 presents summary statistics of regression (30). The bias of the CSV measured with respect to standard asset pricing models is small in magnitude for both weighting schemes (in the order of  $10^{-5}$ , but remain statistically significant. Overall, we can safely consider that the impact of the bias remains immaterial for any practical purposes. Another interesting fact is the sign of the biases. For the equal weighted quantities, the sign of the bias is positive, while the same intercept for the cap-weighted ones is negative. This means that the beta bias dominated the weighting bias for equal-weighted averages in both models. This is consistent with the prediction made by the theoretical analysis regarding the relative impact of the weighting-bias for different weighting schemes. On the other hand, the bias has, in both cases, a higher magnitude when measured with respect to the Fama-French model than when measured with respect to the CAPM<sup>16</sup>.

### 3.2 Comparison with other Measures

In this section we compare the CSV measure to the afore-mentioned, more conventional, measures of idiosyncratic risk (i.e. FF-based, CAPM-based and GS). To obtain these other measures, we need to re-estimate the relevant factor model using a rolling window of one month worth of daily data to allow for time-variation in beta estimates (or total variance variation for the GS). Table 3 presents summary statistics for the monthly time-series of annualized idiosyncratic variances based on 516 observations from January 1964 to December 2006<sup>17</sup>.

On the monthly series, the annualized mean of the equally-weighted CSV, FF-based and CAPM-based measures are 38.42%, 38.32% and 38.79%, respectively, with a difference of less than 0.5%, while the EW GS variance is 34.24%. The standard deviations are

---

<sup>16</sup>The adjusted  $R^2$  of around 99% in this regression should not surprise the reader since these two quantities (the regressor and the CSV) have Pearson correlation coefficients of around 90% as well, as documented in Section 3.2.

<sup>17</sup>In this section of the paper, we start the sample period in January 1964 to allow for direct comparison with Bekaert et al. (2008). In the predictability section, we instead start the sample in July 1963.

8.52%, 8.60%, 8.66% for the CSV, FF-based and CAPM-based measures and 6.99% for the GS measure. For the cap-weighted version, the CSV, FF and CAPM idiosyncratic variance measures have an annualized mean of 8.52%, 7.62%, 8.04% respectively and the GS measure mean is 11.23%. The standard deviations are also closer among CSV, FF and CAPM than with the one of GS. Although GS argue that their measure *fundamentally* constitutes a *measure of idiosyncratic risk*, with the idiosyncratic component accounting for about 85% of the total EW average measure, it is strictly speaking an average of total stock variance. Our measure generates empirical results that are close to idiosyncratic variance measures derived from traditional asset pricing models, suggesting that the assumption about the beta homogeneity does not constitute a major problem to capture idiosyncratic risk. That our measure agrees with previously introduced measures is confirmed by the cross-correlation analysis reported in Table 4.

The smallest correlation coefficient in Table 4 is 56.98% between the  $GS^{CW}$  and  $FF^{EW}$  measure. The highest correlation in the table is 99.93% between the  $CSV^{EW}$  and the  $FF^{EW}$ . Among the CW measures, the highest correlation is 99.49% between CSV and CAPM, followed by 99.19% between CAPM and FF and the lowest between GS and FF with 88.19%. Among the EW measures the cross-correlations with the GS measure are also the lowest.

Table 5 provide mean and standard deviation estimates for the average idiosyncratic variance measures when daily estimates are used. The mean of the equally-weighted CSV is 39.54% while the EW idiosyncratic variance relative to the FF model is 38.33%. For the cap-weighted measures the CSV has slightly higher mean than the FF-based one. In contrast with the monthly frequency, the CSV daily series presents a higher standard deviation with respect to the FF-based average idiosyncratic variance in both cap-weighted and equal-weighted forms. This is not surprising given that the daily CSV measures are *instantaneous*, since they only include information of the cross-section. This contrasts with the FF idiosyncratic variance measure that contains a built-in artificial smoothing element because it requires a window of data to estimate the  $N_t \times NF$  (with NF being the number of factors of the asset pricing model) coefficients of the factor model in addition to the  $N_t$  idiosyncratic variances. Hence, each estimate for idiosyncratic variance relative to a factor model differs from the previous one only by two observations out of approximately 21 trading days in a month (for a monthly rolling window sample), the first and last day of the each sample. In this sense, today's estimate is about 90% yesterday's estimate<sup>18</sup>.

The intuition that the rolling-window methodology inherent in standard measures generates a smoothed estimate of current idiosyncratic risk level is confirmed in Figures 1 and 2 which plots daily CSV and FF-relative idiosyncratic variance, using an equal-weighted

---

<sup>18</sup>This is because  $\frac{21-2}{21} = 90.48\%$

and a cap-weighted average, respectively. It should also be noted that the estimation of the FF-measure is computationally much more expensive than for the CSV measure, which is based on observable quantities.

Table 6 presents cross-correlations for the daily series idiosyncratic variance measures. Although the coefficients are considerably smaller than for the monthly series, the relationship remains strong provided the comparison is done for the same weighting scheme. The difference with the monthly series correlations may again be explained by the presence of the smoothed estimation procedure inherent to the FF-based measure.

### 3.3 Extracting Regimes in Idiosyncratic Risk

Following Bekaert et al. (2008), we fit a Markov regime switching model with a one-lag autocorrelation structure (see Hamilton (1989)). In this model, two regimes are indexed by a discrete state variable,  $s_t$ , which follows a Markov Chain process with constant transition probabilities. Let the current regime be indexed by  $i$  and the past regime by  $j$  and  $x_t$  be the original idiosyncratic variance. In this parsimonious model,  $x_t$  follows an AR(1) model:

$$x_t - \mu_i = \phi(x_{t-1} - \mu_j) + \sigma_i e_t, \quad i, j \in \{1, 2\} \quad (31)$$

The transition probabilities are denoted by  $p = P[s_t = 1 | s_{t-1} = 1]$  and  $q = P[s_t = 2 | s_{t-1} = 2]$ . The model involves a total of 7 parameters,  $\{\mu_1, \mu_2, \sigma_1, \sigma_2, \phi, p, q\}$ .

The estimation results for the monthly series of both  $FF^{CW}$ ,  $CSV^{CW}$ ,  $FF^{EW}$  and  $CSV^{EW}$  are reported in Table 7. For corresponding weighting schemes, the parameters in both regimes are similar between the two measures. For both measures the low-mean low-variance regime presents a higher probability of remaining in the same state.

We then fit the same model to the daily time series and present the parameter estimates in Table 8. It should be stressed that for our CSV measure, the parameter values of average level of idiosyncratic variance  $\mu$  in both regimes are notably close to the values obtained with the monthly series. This result suggests that the process observed at the daily frequency is not just a noisy series but it actually captures the same underlying process observed at the monthly frequency. This contrasts with the FF-based measure, for which the maximum-likelihood estimation procedure could not recognize two regimes when daily data is used, as evidenced by the fact that the parameter values for the mean level of idiosyncratic variance are basically the same for the two regimes. This problem combined with an autocorrelation parameter very close to one is likely caused by the overlapping data problem present in the daily FF measure, which corresponds to the *smoothing* effect mentioned in the previous subsection.

In Figures 3 and 4 we plot the smoothed probability of remaining in state 1 (high-mean

-high variance regime)<sup>19</sup>, as well as the monthly CSV and FF average idiosyncratic variance time series for the CW and EW weighting schemes, respectively. In the legend for the  $x$  axis, we have identified selected months that mark obvious changes in regime dates. At the monthly frequency, our measure and the FF-based appear to be remarkably close for both the equal-weighted and cap-weighted average schemes. Also, we find that the dates of regime changes, marked by the smoothed probability, are the same most of the times for the cap-weighted and the equal-weighted measures<sup>20</sup>. We also find that periods in the higher-mean and higher-variance regime are more persistent for the equally-weighted measure compared to the cap-weighted measure (except during the tech bubble period). Overall, our smoothed probability series resembles closely the one presented in Bekaert et al. (2008) for the cap-weighted FF and Campbell et al. (2001) measures. The small difference might come from the fact that Bekaert et al. (2008) fit a model with two different autocorrelation coefficients (one for each regime) as opposed to one. However, they find the two coefficients to be fundamentally equal in both regimes, which supports using a more parsimonious model.

Finally in Figures 6 and 5 we plot again the smoothed (conditional) probability of remaining in the high-mean high-variance regime together with the NBER recession (or contraction) periods (shaded areas). The peaks in the probability coincides most of the times with the NBER recession periods, which confirms the counter-cyclical property of the CSV measure.

## 4 Idiosyncratic Skewness and Cross-Sectional Third-Order Moments

Recent research have found that idiosyncratic skewness may also have a relationship with expected returns. For instance Boyer et al. (2009) find a significant relationship between expected idiosyncratic skewness and future returns at the stock level. In order to explore this relationship at the aggregate level, we now use an approach similar the one used in section 2.1.

Let us look at the relationship between aggregate idiosyncratic skewness and third order generalizations of our cross-sectional variance measure of equation (4). Consider the standardized  $3th$  central moment or (cross-sectional) skewness measure:

---

<sup>19</sup>Smoothed probabilities are estimations of the transition probability conditional to information up to time  $t$ .

<sup>20</sup>One notable exception is the regime change of 1980 : 05, which is present for the cap-weighted measure and absent for the equally-weighted one.

**Definition (CSS):** The *cross-sectional skewness*, denoted by  $CSS_t$ , is given by

$$CSS_t = \frac{cm_{3t}}{cm_{2t}^{3/2}} = \frac{\frac{1}{N_t} \sum_{i=1}^{N_t} (r_{it} - \bar{r}_t)^3}{\left( \frac{1}{N_t} \sum_{i=1}^{N_t} (r_{it} - \bar{r}_t)^2 \right)^{3/2}}$$

where  $r_{it}$  is the  $i^{th}$  stock return,  $\bar{r}_t$  is the equal-weighted portfolio return,  $cm_3$  is the cross-sectional third central moment, and  $cm_2$  is the cross-sectional variance, hence  $cm_{2t} = CSV_t^{EW}$ . For completeness, we also consider a market-cap weighted version of this measure denoted by  $CSS^{CW}$ .

Denote the idiosyncratic skewness of the  $i^{th}$  stock as

$$sk_{\varepsilon_i}(t) = \frac{m_3(\varepsilon_{it})}{m_2(\varepsilon_{it})^{3/2}} \quad (32)$$

A similar simplification to Assumption 1 but on the third central moment allows us to establish a link between the *CSS* and idiosyncratic skewness in the market.

**Assumption 5**  $\forall i, t$ ,  $\varepsilon_{it}$  is independently distributed in time with homogeneous third central moment across assets

$$E(\varepsilon_{it}^3) \equiv m_3(\varepsilon_t)$$

For illustration purposes, again, let us first look at the simple case with homogeneous betas. In this case and under Assumptions 1 and 5, the following Corollary holds.

**Proposition 3 (CSS as proxy for idiosyncratic skewness):** Under assumptions 1, 2 and 5,

$$CSS_t \xrightarrow{N_t \rightarrow \infty} sk_{\varepsilon}(t) \text{ almost surely} \quad (33)$$

In the single index case where  $F_t = r^{(wt)}$  this result actually follows in an exact manner, meaning  $CSS_t = sk_{\varepsilon}(t)$ . The proof of Corollary 3 goes as follows

**Proof** Using the factor decomposition for returns we have,

$$r_{it} - r_t^{EW} = \varepsilon_{it} - \frac{1}{N_t} \sum_{j=1}^{N_t} \varepsilon_{jt}$$

From the strong law of large numbers, we know that  $\frac{1}{N_t} \sum_{j=1}^{N_t} \varepsilon_{jt} \xrightarrow{N_t \rightarrow \infty} 0$  almost surely. In other words, for large  $N_t$ , we have that  $r_{it} - r_t^{EW} \rightarrow \varepsilon_{it}$ <sup>21</sup>. Replacing this result in the *CSS* definition yields,

$$CSS_t \xrightarrow{N_t \rightarrow \infty} \frac{\frac{1}{N_t} \sum_{i=1}^{N_t} \varepsilon_{it}^3}{\left( \frac{1}{N_t} \sum_{i=1}^{N_t} \varepsilon_{it}^2 \right)^{3/2}}$$

---

<sup>21</sup>When  $F_t = r^{EW}$  this result actually follows in an exact manner:  $r_{it} - r_t^{EW} = \varepsilon_{it}$

From Assumptions 1 and 5 and the definition of skewness it follows

$$CSS_t \xrightarrow{N_t \rightarrow \infty} sk_\varepsilon(t) \text{ almost surely} \quad \blacksquare$$

Corollary 3 is an intuitive argument that draws a direct relationship between the CSS and idiosyncratic skewness, however it holds under the simplistic assumption of homogenous betas across assets. Let us now look at the impact of the beta's cross-sectional distribution but relax as well the homogeneous central moments Assumptions 1 and 5. Consider the factor decomposition for the case where  $F_t = r^{EW}$  (sometimes denoted as  $\bar{r}_t$ )

$$r_{it} - r_t^{EW} = (\beta_{it} - 1) r_t^{EW} + \varepsilon_{it} \quad (34)$$

Replacing (34) in the CSS definition,

$$CSS_t = \frac{cm_{3t}}{cm_{2t}^{3/2}} = \frac{\frac{1}{N_t} \sum_{i=1}^{N_t} ((\beta_{it} - 1) r_t^{EW} + \varepsilon_{it})^3}{\left( \frac{1}{N_t} \sum_{i=1}^{N_t} ((\beta_{it} - 1) r_t^{EW} + \varepsilon_{it})^2 \right)^{3/2}} \quad (35)$$

The numerator in (35), the third central moment  $cm_{3t}$  expands to,

$$cm_{3t} = \bar{r}_t^3 \frac{1}{N_t} \sum_{i=1}^{N_t} (\beta_{it} - 1)^3 + 3\bar{r}_t^2 \frac{1}{N_t} \sum_{i=1}^{N_t} (\beta_{it} - 1)^2 \varepsilon_{it} + 3\bar{r}_t \frac{1}{N_t} \sum_{i=1}^{N_t} (\beta_{it} - 1) \varepsilon_{it}^2 + \frac{1}{N_t} \sum_{i=1}^{N_t} \varepsilon_{it}^3$$

Taking the expectation of  $cm_{3t}$  above, denoting  $m_{3t}^\beta$  the third central moment of betas' cross-sectional distribution and noting that  $E[\varepsilon_{it}] = 0$  yields,

$$\begin{aligned} E[cm_{3t}] &= E\left[\bar{r}_t^3 m_{3t}^\beta\right] + 3\bar{r}_t \frac{1}{N_t} \sum_{i=1}^{N_t} (\beta_{it} - 1) \sigma_{\varepsilon_{it}}^2 + \frac{1}{N_t} \sum_{i=1}^{N_t} m_3(\varepsilon_{it}) \\ &= E\left[\bar{r}_t^3 m_{3t}^\beta\right] + 3\bar{r}_t \left( \frac{1}{N_t} \sum_{i=1}^{N_t} \beta_{it} \sigma_{\varepsilon_{it}}^2 - \overline{\sigma_{\varepsilon_{it}}^2} \right) + \frac{1}{N_t} \sum_{i=1}^{N_t} m_3(\varepsilon_{it}) \end{aligned}$$

The second term is likely to be small if the dispersion of betas and/or the heterogeneity of idiosyncratic variances is not too large. In fact the term disappears under Assumption 1 (homogenous idiosyncratic variances) although allowing for heterogeneous third central moment yields,

$$E[cm_{3t}] = E\left[\bar{r}_t^3 m_{3t}^\beta\right] + \frac{1}{N_t} \sum_{i=1}^{N_t} m_3(\varepsilon_{it}) \quad (36)$$

Equation (36) suggests that the cross-sectional third central moment is a biased proxy of average idiosyncratic 3th central moment. The size of its bias (the first term in (36)) would be small for an approximately symmetric beta's cross-sectional distribution.

Regarding the denominator in (35), from (13) in section 2.2 we know that

$$CSV_t^{EW} = cm_{2t} = \bar{r}_t^2 \frac{1}{N_t} \sum_{i=1}^{N_t} (\beta_{it} - 1)^2 + \frac{1}{N_t} \sum_{i=1}^{N_t} \varepsilon_{it}^2 + 2\bar{r}_t \frac{1}{N_t} \sum_{i=1}^{N_t} (\beta_i - 1) \varepsilon_{it}$$

Which after taking the expectation and assuming  $\bar{r}_t$  orthogonal to  $\beta_t$ ,

$$E[CSV_t^{EW}] = E[cm_{2t}] = E[\bar{r}_t^2 CSV_t^\beta] + \frac{1}{N_t} \sum_{i=1}^{N_t} \sigma_{\varepsilon_{it}}^2 \quad (37)$$

Provided that the magnitude of the biases in equations (36) and (37) be negligible, our results suggests that the expected value of cross-sectional third and second central moments are good proxies for their average idiosyncratic counterparts. However, this does not directly implies that their ratio (taking  $cm_2^{3/2}$ ) would be a good proxy for the average idiosyncratic skewness, as one might be tempted to think by looking at equation (35). The reasons for this are two properties of the expectation operator: First,  $E[cm_2^{3/2}]$  is not equal to  $E[cm_2]^{3/2}$ , although it is well known that for the square root approximation (standard deviation) this does not introduce a meaningful bias for a large sample, the third power is more likely to have a sensible impact. Second, the expectation of a ratio of two variables  $x$  and  $y$ , is not the ratio of the expectations, but instead<sup>22</sup>

$$E\left[\frac{x}{y}\right] = \frac{E[x]}{E[y]} \left[1 + \frac{\sigma_y}{\mu_y^2} - \frac{\sigma_{xy}}{\mu_x \mu_y}\right] \quad (38)$$

In different contexts, some authors<sup>23</sup> have used “the simplifying assumption that the expected value of a ratio is approximately the ratio of the expected values”, which, on the one hand would imply from equation (35) that

$$E[CSS_t] = E\left[\frac{cm_3(t)}{cm_2(t)^{3/2}}\right] \approx \frac{E[cm_3(t)]}{E[cm_2(t)^{3/2}]} \quad (39)$$

The expected value of a ratio is greater than or equal to the ratio of the expected values, hence, the right side of (39) is a lower limit. Using the further approximation  $E[cm_2^{3/2}] \approx E[cm_2]^{3/2}$ , could probably compensate this underestimation<sup>24</sup> given  $E[cm_2]^{3/2}$  might be lower than  $E[cm_2^{3/2}]$ . On the other hand, from results (36) and (37) and ignoring the corresponding biases, we have

$$\frac{E[cm_3(t)]}{E[cm_2(t)^{3/2}]} \approx \frac{\frac{1}{N_t} \sum_i^{N_t} m_{3t}(\varepsilon_i)}{\left(\frac{1}{N_t} \sum_i^{N_t} m_{2t}(\varepsilon_i)\right)^{3/2}} \quad (40)$$

<sup>22</sup>Taken from equation (16) of Durmaz and Baydere (2004)

<sup>23</sup>For instance Brennan (2005) and Azzam et al. (1988) use this approximation

<sup>24</sup>This would be the case if for instance the correlation between  $cm_2^2$  and  $cm_2$  is positive, since  $E[x^3] = E[x]^3 + E[x]Var[x] + Cov(x^2, x)$

together with (39) would imply that the CSS could be a rough approximation for average idiosyncratic skewness<sup>25</sup>.

$$E [CSS_t] \approx \frac{1}{N_t} \sum_{i=1}^{N_t} sk_{\varepsilon_i}(t) \quad (41)$$

The relationship outlined above together with Corollary 3 provides the motivation to look at *some* measure of cross sectional skewness as a proxy for aggregate idiosyncratic skewness. The main advantage of using an estimate of cross-sectional skewness as proxy for idiosyncratic skewness as opposed to aggregating individual skewness estimates is twofold. First, we do not need to choose a particular factor model and hence we avoid any possible (betas) estimation issues (which for individual stock returns are rather common). Second, we have observable series even at the daily frequency, which could not be obtained without using our cross-sectional approximation. The *formal* motivation for using third order cross-sectional moments to approximate *aggregate* idiosyncratic skewness (and average idiosyncratic third order central moment) provided in this section, which further exploits the natural link of the cross-sectional distribution with idiosyncratic risk, is entirely new to our knowledge.

In terms of empirical estimation, we follow Kim and White (2004) who argue in favor of the the quantile-based coefficient of skewness (see Bowley (1920)), generalized by Hinkley (1975), as a robust measure of skewness:

$$RCS = \frac{F^{-1}(1 - \alpha_1) + F^{-1}(\alpha_1) - 2Q_2}{F^{-1}(1 - \alpha_1) + F^{-1}(\alpha_1)} \quad (42)$$

for any  $\alpha_1$  between 0 and 0.5 and  $Q_2 = F^{-1}(0.5)$ . The Bowley coefficient of skewness is a special case of Hinkley's coefficient when  $\alpha_1 = 0.25$  and satisfies the Groeneveld and Meeden (1984)'s properties for reasonable skewness coefficients. It has upper and lower bounds  $\{-1, 1\}$ .

Table 9 presents summary statistics for the distribution of the cross-sectional skewness measures mentioned above. The sample period is 1963:07-2006:12. One can see that the distribution of the non-robust measures are highly dispersed and unlikely to convey tractable information. On the other hand the robust measure of skewness is (by construction) much better behaved.

In the next section, we look at the predictability power of idiosyncratic risk on the market return using the cross-sectional CSV before adding as a predictive variable the robust measure of skewness.

---

<sup>25</sup>At the stock level, Zhang (2009) finds an empirical relationship between peer group cross-sectional skewness and individual stock skewness and Kapadia (2009) uses cross-sectional skewness to explain the puzzling finding in Ang et al. (2006) that stocks with high idiosyncratic volatility have low subsequent returns.



## 5 New Evidence on the Predictability of the Market Return

There is an ongoing debate on the predictive power of average idiosyncratic variance over average (or aggregate) stock market returns. Goyal and Santa-Clara (2003) find a significantly positive relationship between the equal-weighted average idiosyncratic stock variance and the cap-weighted portfolio returns for the period 1963:07 to 1999:12. They find that their measure of average (*mostly*) idiosyncratic variance has a significant relation with next month's return on the cap-weighted portfolio. The regression in GS is as follows:

$$r_{t+1}^{CW} = \alpha + \beta \nu_t^{EW} + \varepsilon_{t+1}, \quad (43)$$

where  $\nu_t^{EW}$  corresponds to  $GS_t^{EW}$ . In a subsequent analysis, Bali et al. (2005) argue that this relationship disappeared for the extended sample 1963:07 to 2001:12, and attribute the relationship observed in GS to high-tech-bubble-type stocks (i.e., stocks traded on the NASDAQ) and a liquidity premium. In a similar way, Wei and Zhang (2005) find that the significance of the relationship found by GS disappeared for their sample 1963:07 to 2002:12 and argue that the presumably temporary result of GS was driven mainly by the data in the 1990s. Wei and Zhang (2005) criticize the fact that GS looked at the relationship between an equally-weighted average stock variance and the return on a cap-weighted average stock return, as opposed to an equally-weighted portfolio return. Moreover, both Bali et al. (2005) and Wei and Zhang (2005) find no significant relationship between the cap-weighted measures and the cap-weighted portfolio return in all three sample periods (ending in 1999, 2001 and 2002, respectively). Also, Wei and Zhang (2005) find a significantly positive relation between the equal-weighted GS measure and the equal-weighted market return for the initial sample and a significant positive relation between the equal-weighted cross-sectional variance (i.e. the CSV calculated using monthly returns) and the equal-weighted portfolio return for most sample periods they look at. Nevertheless Wei and Zhang (2005) underweight this last finding and conclude that the relationship is insignificant upon their findings using the competing measures that average idiosyncratic risk.

### 5.1 Monthly Evidence

In this section we confirm existing results, and extend them in a number of dimensions, including a longer sample period. Table 10 presents the predictability regression of equal-weighted variance measures on the cap-weighted return as in Goyal and Santa-Clara (2003) and Bali et al. (2005) for their sample periods and the extended sample up to 2006:12. The regression is as in equation 43, where  $\nu_t^{EW}$  corresponds to the EW CAPM-based measure

and the CSV <sup>26</sup>. For comparison purposes we start the sample period in this section in 1963:07, as in Goyal and Santa-Clara (2003), Bali et al. (2005) and Wei and Zhang (2005).

For the monthly series, we confirm that there is a significant positive relationship in the first sample, and also that it weakens for the subsequent extended samples<sup>27</sup>. The t-stat of the  $\beta$  coefficient of both CSV and the CAPM-based measures goes from 3.2 for the first sample period down to 1.6 for the largest sample. Consequently, the adjusted  $R^2$  goes from 2.11% down to 0.49%. This result confirms the finding of Bali et al. (2005) and Wei and Zhang (2005) for the further extended sample. In section 6 we propose an explanation for this puzzling finding. However, as argued in Wei and Zhang (2005), it seems more natural to look at the relationship between similar weighting schemes in both average measures (return and variance).

In Table 11 we present the results of the regression between the equal-weighted average return with the lagged equal-weighted average idiosyncratic variance measures given by:

$$r_{t+1}^{EW} = \alpha + \beta \nu_t^{EW} + \varepsilon_{t+1} \quad (44)$$

where  $\nu_t^{EW}$  is taken as the CAPM-based average idiosyncratic variance or as the CSV measure. In contrast with the former regression, the relationship is found to be significantly positive for the three sample periods (for both measures).

In Table 12 we present the results for the three sample periods of the one-month-ahead predictive regression of the cap-weighted market portfolio using the cap-weighted idiosyncratic variance return as a predictor. In this case, the beta of the idiosyncratic variance is significant for none of the three sample periods. This result confirms the findings of Bali et al. (2005) and Wei and Zhang (2005) for the extended sample.

## 5.2 New Predictability Evidence at Daily Frequency

Prevailing measures in the literature need a sample of past data to estimate additional parameters, constraining existing evidence to the monthly estimations. Fu (2009) find that high idiosyncratic volatilities of individual stocks are contemporaneous with high returns, which tend to reverse in the following month. Huang et al. (2009) find that the negative relationship between idiosyncratic variance and expected returns at the stock level found in Ang et al. (2006) and Ang et al. (2008) is positive after controlling for the return reversals. This provides additional motivation to look at the predictability relation at a higher frequency than the monthly basis. Using the CSV as a proxy for aggregate idiosyncratic

<sup>26</sup>As explained before, the monthly CSV is the average of its daily estimations during the month.

<sup>27</sup>We found a similar result using the GS measure of equal-weighted average variance. We do not present these regression results for the sake of brevity given that they generate a similar result, which has also been confirmed in Bali et al. (2005) and Wei and Zhang (2005).

variance allows us to check this relationship at the aggregate (market) level in a more direct way (without having to control for reversals). Taking advantage of the instantaneous nature of the CSV, we run the same predictability regression (44) on the one-day-ahead portfolio return using the average idiosyncratic variance.

Table 13 shows that at a daily basis, this relationship is much stronger, with (Newey-West autocorrelation corrected) t-stats of coefficient for the the average idiosyncratic variance across the three samples between 4 and 4.9. However, the  $R^2$  deteriorates with respect to the first sample period, going from 0.93% of the sample ending in December 1999 down to 0.59% for the most updated data set. Table 14 presents the results for the one-day-ahead predicting regression on the cap-weighted pairs (CSV and market return) and find the relation also to be positive and significant, but with a much more obvious deterioration of the t-stat of the cap-weighted idiosyncratic variance coefficient, going from about 5.7 in the first sample down to 2.08. For this reason and for brevity, we now focus on the relationship between aggregate idiosyncratic risk and the equal weighted market return<sup>28</sup>.

### 5.3 Robustness

In order to check wether the relationship between the market portfolio expected return and the aggregate level of idiosyncratic variance that we document at the monthly and daily frequency, is robust to the inclusion of the variance of the market portfolio, we run the following joint regression,

$$r_{t+1}^{EW} = \alpha + \beta CSV_t + \vartheta Var(r_t^{EW}) + \epsilon_{t+1}. \quad (45)$$

We also run the univariate regression

$$r_{t+1}^{EW} = \alpha + \vartheta Var(r_t^{EW}) + \epsilon_{t+1}. \quad (46)$$

For the monthly estimations of  $Var(r_t^{EW})$  we use the realized sample variance over the month (from daily returns). For daily estimations we fitted an AR(1)-EGARCH(1,1) model on the overall sample<sup>29</sup>. Table 15 presents the regression results at the monthly and daily frequency of both (45) and (46). On the latter univariate regression, the variance of the market portfolio presents a positive non-significant relationship (with  $\vartheta = 1.5$ ) with the market expected return at the monthly basis. This relationship is negative and non-significant at the daily frequency with a coefficient value  $\vartheta$  of  $-1.5$ .

---

<sup>28</sup>The corresponding results using a market cap-weighted scheme can be obtained from the authors upon request.

<sup>29</sup>Using the overall sample to estimate the parameters would only give the portfolio variance an advantage to predict future returns. However, from the results we see that even using this forward looking estimates for  $Var(r_t^{EW})$  the significance of the CSV remains strong.

Consistent with results in Goyal and Santa-Clara (2003), on the joint regression, the coefficient of  $Var(r_t^{EW})$  is negative and non-significant at monthly frequency. On the daily horizon regression, the coefficient was found still negative and (marginally) significant. The significance of the  $CSV$  coefficient, at both monthly and daily basis, improved slightly after the inclusion of the portfolio variance.

Finally, through the following regression we explore *i*) whether the predictability power is the same for both, the returns to the left and right of the cross-sectional distribution, *ii*) whether the relationship is driven by one of the sides and *iii*) whether both sides would have the same sign on their coefficient. In order to do this we define the  $CSV_t^+$  as the cross-sectional variance of the returns to the right of the cross-sectional distribution (right meaning that include all stocks such that  $r_{it} > r_t^{EW}$ ) and conversely define the  $CSV_t^-$  as the cross-sectional variance of the returns to the left of the cross-sectional distribution. Then we run the following regression.

$$r_{t+1}^{EW} = \alpha + \beta^+ CSV_t^+ + \beta^- CSV_t^- + \epsilon_{t+1} \quad (47)$$

Table 16 presents the results of regression (47) for daily and monthly estimates which has a couple of interesting facts. First, splitting the  $CSV$  into right and left sides of the cross-sectional distribution, made the adjusted  $R^2$  of the predictive regression jump from 0.8% to 1.17% on monthly data and from 0.6% to 1.36% on daily data. Second, there is an asymmetric relationship between the  $CSV$  of the returns to the right and left of the cross-sectional distribution and the expected market return: the coefficient of the  $CSV_t^+$  is positive while the one of  $CSV_t^-$  is negative in both daily and monthly regressions. However, the coefficients (of both right and left  $CSV$ s) are significant only on the daily regression.

These facts suggest that a measure of asymmetry of the cross-sectional distribution would be pertinent in the context of exploring the relationship between market expected returns and aggregate idiosyncratic risk.

## 5.4 Improved Predictability Evidence with Cross-Sectional Skewness

In order to further test for the predictive power of aggregate idiosyncratic risk in a broader sense, we now integrate cross-sectional skewness. To the best of our knowledge, this additional factor, which was provided here as a natural extension of the  $CSV$  in section 4, is entirely new in this context<sup>30</sup>. Using the same setting we run the following one-day-ahead

---

<sup>30</sup>At the stock level, Kapadia (2009) uses cross-sectional skewness to explain the puzzling finding in Ang et al. (2006) that stocks with high idiosyncratic volatility have low subsequent returns.

and one-month-ahead regression

$$r_{t+1}^{EW} = \alpha + \beta_1 CSV_t^{EW} + \beta_2 CSS_t + \varepsilon_{t+1} \quad (48)$$

where  $\beta_2$  is the OLS coefficient of the *Robust* Cross-sectional Skewness (estimated as given in equation (42)) and  $\beta_1$  the coefficient associated with the *CSV*.

The inclusion of the robust measure of cross-sectional skewness has a remarkable impact in the predictive power of idiosyncratic risk on the market portfolio return. Table 17 presents the results of the one-day-ahead regression as described in equation (48). The combined effect of the  $CSV^{EW}$  and the robust measure of cross-sectional skewness on the next period equal-weighted return generates an adjusted  $R^2$  of 7.4% on the sample ending in 1999 and of 5.8% on the most updated series (ending in December 2006). The coefficient of the CSV remained strong (in fact for the first sample t-stat is slightly higher) after introducing the cross-sectional skewness into the predictive regression, with Newey-West t-stats between 4 and 5 for the three samples considered. The significance of the second regressor (the robust cross-sectional skewness) is remarkably high, with Newey-West autocorrelation corrected t-stats between 19.8 and 20. The coefficient values are stable across sample periods,  $\beta_1$  took values of between 0.4 and 0.52 while  $\beta_2$  was 0.4% for the three samples.

The monthly regressions, displayed in Table 18 present a similar pattern. The adjusted  $R^2$  are again (comparatively) high, between 4.6% and 5% for the three periods. The  $CSV^{EW}$  coefficient t-stat remains very significant and positive, with Newey-West autocorrelation corrected of around 2.6 for all three samples, while the robust c-skewness coefficient's t-stats are of 4.12 (for the sample ending in Dec. 99), 4.28 (Dec. 01) and 4.45 for the whole sample (up to Dec. 06). The coefficient values are also stable across sample periods on the monthly regression with  $\beta_1$  taking values between 0.25 and 0.27 and  $\beta_2$  between 7.3% and 7.8% for the three samples <sup>31</sup>.

## 6 Interpretation

In what follows we offer some interpretation of our results and of previous evidence regarding the relation between average idiosyncratic risk and aggregate market return. We provide an explanation for previous puzzling findings and regroup them with the positive evidence on the relevance of idiosyncratic risk.

---

<sup>31</sup>In unreported results, we have also run regression (48) using the conventional measure of skewness as the second explanatory variable, but it did not generate any better results. This last negative result is not very surprising given the high instability of the measures, which is illustrated in Section 4 and more specifically in Table 9.

## 6.1 On the Market-Cap Puzzle

After seeing the presented evidence on the predictability of idiosyncratic risk on average market return, a natural question to ask would be: why does this relationship differ across weighting schemes in different sample periods?

There are multiple reasons for which average idiosyncratic risk should be related to average returns, or in other words, there is no unique reason for this relation to exist. It is precisely the nature of idiosyncratic risk that makes this relationship heterogeneous. By construction, idiosyncratic risk represents the residual part of the return that is not captured by the explicit factor(s) of a particular asset pricing model. As such, the nature and importance of omitted factors may vary across periods. Interestingly, this paper and previous research suggest that most results on the properties and dynamics of idiosyncratic risk are robust across different measures of average idiosyncratic variance, given a particular defined weighting scheme is chosen.

As mentioned in the introduction, Campbell et al. (2001)(2001) and Goyal and Santa-Clara (2003) recall some of the possible explanations for the relevance of idiosyncratic risk. From this perspective, one may see findings on Bali et al. (2005) and Wei and Zhang (2005) supporting an additional explanation by stressing the fact that the particular dot.com bubble companies played an important role in the relation with the average market-capitalization return during the end of the 1990s, which (obviously) weakened after the burst of the bubble. In this sense Bali et al. (2005) argues that the relation between (equal-weighted) average idiosyncratic risk and the market-cap weighted index on the sample ending in 1999:12 is mostly driven by small stocks traded in the NASDAQ.

Our results confirmed findings on Bali et al. (2005) and Wei and Zhang (2005). It is well documented that during the dot.com bubble “abnormal” returns inflated the valuation of several small NASDAQ companies. The strongest omitted factors of that period (call it the irrational.com factor) partially captured by the equally weighted idiosyncratic variance, started to be increasingly represented in the market-cap index, due to the suddenly-higher market capitalization of precisely the group of companies carrying this temporarily strong omitted factor. The posterior reversal of the situation (i.e. the burst of the bubble) explains the sharp fade in the relationship between the average idiosyncratic variance and the market-cap portfolio, precisely due to the posterior sudden deterioration of the market capitalization of most stocks carrying this irrational.com factor, and hence notably reducing their representability in the market-capitalization index.

## 6.2 On the Equal-Weighted Relation: from Monthly and Daily Evidence

Some intuition behind the far more robust relationship between the equally-weighted average idiosyncratic variance and the equally weighted portfolio comes precisely from the logic of standard asset pricing. According to CAPM, only systematic risk (i.e. covariance with the market index) should explain future returns. However, if during a certain period of time there exists an anomaly of any kind (e.g. value, size, irrational.com), which presumably does not automatically affect the current market capitalization of the companies carrying that factor, then this omitted factor is more likely to explain the returns of a portfolio where all kinds of firms are represented in a similar manner as opposed to a portfolio where big companies are proportionally better represented than smaller ones. This is of course because the omitted factors are not necessarily proportionally affecting firms according to their current market capitalization. Nevertheless, as we have seen, this situation may occur, like during the dot.com bubble, but this kind of situation will only hold during the time when the idiosyncratic factors (or anomalies) are strongly contemporaneously reflected by the market capitalization of firms (which in the case of the irrational.com anomaly was indeed contemporaneous and approximately proportionally reflected).

One fair remark upon the results of the predictability regressions, is that at a higher frequency (i.e. daily), the relationship seems to be comparatively stronger than at the monthly basis. This should not come as a surprise given the evidence presented by Fu (2009) at the stock level, who finds that high idiosyncratic volatilities of individual stocks are contemporaneous with high returns, which tend to reverse in the following month.

Our measures of idiosyncratic risk, i.e. CSV and its higher-moments extension, are basically a dynamic picture of the cross-sectional distribution of realized returns. They allow us to extract information about the diversity of the companies with respect to each other by subtracting the effect of the aggregate return. As we have formally shown, they are natural measures for idiosyncratic risk, which proxy for any (temporary or permanent) factors omitted in standard asset pricing models. We are inclined to privilege an equal-weighting of firm observations into the global measure of average idiosyncratic variance, since we find no intuitive arguments, statistical advantage nor empirical motivation for using a market-capitalization weighted measure.

## 7 Conclusion

In this paper we formally introduce a measure of aggregate idiosyncratic risk that has the distinct advantage of being readily observable, with no need to estimate other parameters. It is an *instantaneous* cross-sectional measure of average idiosyncratic variance, available at

any given data frequency and that generalizes to higher moments.

We extensively show how this measure is related to previous proxies of idiosyncratic variance, such as the Goyal and Santa-Clara (2003) measure and measures relative to the classic Fama and French (1993) and CAPM models, which have been previously shown to be very close to the Campbell et al. (2001) (2001) proxy as well. We confirm previous findings of Goyal and Santa-Clara (2003), Bali et al. (2005) and Wei and Zhang (2005) on the monthly predictability regressions for the extended sample period using our cross-sectional measure and more standard measures of idiosyncratic variance. We find that the results are robust across these measures. Thanks to the instantaneous nature of our measure, we are able to extend to daily data the evidence on the predictability power of idiosyncratic volatility on the future market portfolio return. Furthermore, we find that the inclusion of a robust measure of cross-sectional skewness into the predictability regression strongly increases the explanatory power of idiosyncratic risk on the equal-weighted portfolio return for both, daily and monthly horizons.

We provide a statistical argument to support the choice of an equally-weighted measure of average idiosyncratic variance as opposed to a market-cap weighted and explain why both empirically and theoretically such a measure should forecast better the equal-weighted market return.



## References

- Ang, A., R. Hodrick, Y. Xing, and X. Zhang (2006). The cross-section of volatility and expected returns. *Journal of Finance* 61, 259–299.
- Ang, A., R. Hodrick, Y. Xing, and X. Zhang (2008). High idiosyncratic volatility and low returns: international and further US evidence. *Journal of Financial Economics*.
- Azzam, S., J. Keele, and M. Nielsen (1988). Expectations of heritability estimates for non-return rate of bulls and conception rate of cows. *Journal of Animal Science* 66(11), 2767.
- Bali, T., N. Cakici, X. Yan, and Z. Zhang (2005). Does Idiosyncratic Risk Really Matter? *The Journal of Finance* 60(2), 905–929.
- Barberis, N., M. Huang, and T. Santos (2001). Prospect Theory and Asset Prices. *Quarterly Journal of Economics* 116(1), 1–53.
- Bekaert, G., R. J. Hodrick, and X. Zhang (2008). Is There a Trend in Idiosyncratic Volatility? *SSRN eLibrary*.
- Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of political economy* 81(3), 637.
- Bowley, A. (1920). Elements of Statistics. *New York: Charles Scribner's Sons*.
- Boyer, B., T. Mitton, and K. Vorkink (2009). Expected idiosyncratic skewness. *Review of Financial Studies*.
- Brandt, M., A. Brav, J. Graham, and A. Kumar (2009). Forthcoming. The Idiosyncratic Volatility Puzzle: Time Trend or Speculative Episodes. *Review of Financial Studies*.
- Brennan, R. (2005). Some Test Theory for the Reliability of Individual Profiles.
- Campbell, J., M. Lettau, B. Malkiel, and Y. Xu (2001). Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk. *The Journal of Finance* 56(1), 1–43.
- Constantinides, G. and D. Duffie (1996). Asset pricing with heterogeneous consumers. *Journal of Political Economy* 104(2), 219.
- DiBartolomeo, D. (2006). Applications of Portfolio Variety.

- Durmaz, O. and S. Baydere (2004). Impact of In-Network Buffering on the Reliability of Sensor Networks. In *Proceedings of the Second International Workshop on Sensor and Actor Network Protocols and Applications (SANPA 2004)*. Boston, Massachusetts, USA. Citeseer.
- Fama, E. and K. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33(1), 3–56.
- French, K. R., G. W. Schwert, and R. F. Stambaugh (1987, September). Expected stock returns and volatility. *Journal of Financial Economics* 19(1), 3–29.
- Fu, F. (2009). Idiosyncratic risk and the cross-section of expected stock returns. *Journal of Financial Economics* 91(1), 24–37.
- Goyal, A. and P. Santa-Clara (2003). Idiosyncratic Risk Matters! *The Journal of Finance* 58(3), 975–1008.
- Groeneveld, R. and G. Meeden (1984). Measuring Skewness and Kurtosis. *The Statistician* 33, 391–399.
- Guo, H. and R. Savickas (2008). Average Idiosyncratic Volatility In G7 Countries. *The Review of Financial Studies* 21(3), 1259–1296.
- Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica* 57(2), 357–384.
- Harvey, C. R. and A. Siddique (2000). Conditional skewness in asset pricing tests. *Journal of Finance* 55, 1263–1295.
- Hinkley, D. (1975). On Power Transformations to Symmetry. *Biometrika* 62, 101–111.
- Huang, W., Q. Liu, S. Rhee, and L. Zhang (2009). Return reversals, idiosyncratic risk, and expected returns. *Review of Financial Studies*.
- Kapadia, N. (2009). The next Microsoft ? Skewness, idiosyncratic volatility, and expected returns. *unpublished paper*.
- Kim, T. and H. White (2004). On more robust estimation of skewness and kurtosis. *Finance Research Letters* 1(1), 56–73.
- Levy, H. (1978). Equilibrium in an Imperfect Market: A Constraint on the Number of Securities in the Portfolio. *American Economic Review* 68(4), 643–658.

- Malkiel, B. G. and Y. Xu (2002). Idiosyncratic risk and security returns.
- Markowitz, H. M. (1990). Risk Adjustment. *Journal of Accounting, Auditing and Finance*.
- Merton, R. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of finance*, 449–470.
- Merton, R. (1987). A simple model of capital market equilibrium with incomplete information. *Journal of Finance* 42, 483–510.
- Sharpe, W. (1963). A simplified model for portfolio analysis. *Management science*, 277–293.
- Wei, S. and C. Zhang (2005). Idiosyncratic Risk Does Not Matter: A Re-examination of the Relationship between Average Returns and Average Volatilities. *Journal of Banking and Finance*, Vol. 29, pp. 603-621, 2005.
- Wei, S. and C. Zhang (2006). Why Did Individual Stocks Become More Volatile? *The Journal of Business* 79(1), 259–292.
- Zhang, Y. (2009). Individual skewness and the cross-section of average stock returns. *unpublished paper, Yale University*.

# A Proof of Proposition 1

**Proof** Consider the factor model decomposition

$$r_t^{(w_t)} = \sum_{i=1}^{N_t} w_{it} \beta_{it} F_t + \sum_{i=1}^{N_t} w_{it} \varepsilon_{it}$$

and

$$r_{it} - r_t^{(w_t)} = \left( \beta_{it} - \sum_{j=1}^{N_t} w_{jt} \beta_{jt} \right) F_t + \varepsilon_{it} - \sum_{j=1}^{N_t} w_{jt} \varepsilon_{jt}$$

Using homogeneous betas as in Assumption 2 we have

$$r_{it} - r_t^{(w_t)} = \varepsilon_{it} - \sum_{j=1}^{N_t} w_{jt} \varepsilon_{jt} \tag{49}$$

In a recent paper, Cuzick (1995) proved the Marcinkiewicz–Zygmund strong law of large numbers for weighted sums of i.i.d. variables

$$\frac{1}{N} \sum_{i=1}^N a_{Ni} X_i \longrightarrow 0 \text{ almost surely} \tag{50}$$

when  $\{X, X_N, N \geq 1\}$  is a sequence of i.i.d. random variables with  $EX = 0$  and  $E|X| < \infty$  and  $\{a_{Ni}, 1 \leq i \leq N, N \geq 1\}$  is an array of constants uniformly bounded satisfying<sup>32</sup>

$$\sup |a_{Nit}| < \infty. \tag{51}$$

We want to rewrite expression (50) taking  $a_{Nit} = N_t w_{it}$  and replacing  $X_i$  with  $\varepsilon_i$ . For result (50) to hold  $a_{Nit}$  needs to be uniformly bounded and to satisfy condition (51).

We restrict our attention to non-trivial weighting schemes, ruling out the situation such that the index is composed by a single stock. Please note that this condition together with the fact that  $\sum_i w_{it} = 1$  implies  $N_t > 1$  and also restrict the weights to be (strictly) positive at every given point in time. Hence, a weighting scheme  $(w_t)$ , is defined as a vector process which satisfies  $0 < w_{it} < 1 \forall i, t$ . This condition seems reasonable since our focus is to measure idiosyncratic risk in the market.

By definition, the weighting scheme  $w_{it}$  and  $a_{Nit}$  is uniformly bounded by  $N_t$  and the following condition holds,

$$0 < w_{it} < 1 \forall i, t \tag{52}$$

---

<sup>32</sup>See Theorem 1.1, particular case of Cuzick (1995).

Multiplying by  $N_t$ , we get

$$\begin{aligned} 0 &< N_t w_{it} < N_t \\ 0 &< N_t w_{it} < \infty \\ 0 &< a_{Nit} < \infty \\ |a_{Nit}| &< \infty \quad \forall i, t \end{aligned}$$

Which implies that condition (51) holds. Thus, for a positive weighting scheme from the strong law of large numbers for weighted sums of i.i.d. variables, it follows that

$$\sum_{i=1}^{N_t} w_{it} \varepsilon_{it} \xrightarrow[N_t \rightarrow \infty]{} 0 \text{ a.s.},$$

which corresponds to the second term at the right hand side of equation (49). Using this result in the definition of the CSV we have that

$$CSV_t^{(w_t)} = \sum_{i=1}^{N_t} w_{it} \left( r_{it} - r_t^{(w_t)} \right)^2 \xrightarrow[N_t \rightarrow \infty]{} \sum_{i=1}^{N_t} w_{it} \varepsilon_{it}^2. \quad (53)$$

Finally, for a diverse weighting scheme,  $w_t$ , and i.i.d.  $\varepsilon_i$ , the approximation  $\sum_{i=1}^{N_t} w_{it} \varepsilon_i^2(t) \simeq \sigma_\varepsilon^2(t)$ , follows from the definition  $E[\varepsilon^2] \equiv \sigma_\varepsilon^2$ . For a positive weighting scheme, the weights can be interpreted as probabilities on the expectation, which implies

$$CSV_t^{(w_t)} \xrightarrow[N_t \rightarrow \infty]{} \sum_{i=1}^{N_t} w_{it} \varepsilon_{it}^2 \simeq \sigma_\varepsilon^2(t) \text{ almost surely} \quad \blacksquare \quad (54)$$

A priori, from a purely statistical standpoint, it is more natural to give equal probabilities to every observation in the sample (which will be the case for the equal-weighted CSV), but sometimes a different type of weighting scheme is of interest, i.e. capitalization weighted scheme.

## A.1 Equally-Weighted CSV and Idiosyncratic Variance

Let the return on the equal-weighted market portfolio be:

$$r_t^{EW} = \frac{1}{N_t} \sum_{i=1}^{N_t} r_{it} \quad (55)$$

where  $N_t$  is the number of stocks in the universe at the given month to which day  $t$  belongs.

Using the factor model decomposition, we have that:

$$r_t^{EW} = \frac{1}{N_t} \sum_{i=1}^{N_t} \beta_{it} F_t + \frac{1}{N_t} \sum_{i=1}^{N_t} \varepsilon_{it}$$

and

$$r_{it} - r_t^{EW} = \left( \beta_{it} - \frac{\sum_{j=1}^{N_t} \beta_{jt}}{N_t} \right) F_t + \varepsilon_{it} - \frac{1}{N_t} \sum_{j=1}^{N_t} \varepsilon_{jt} = (\beta_{it} - 1) F_t + \varepsilon_{it} - \frac{1}{N_t} \sum_{j=1}^{N_t} \varepsilon_{jt}$$

Assume now for simplicity that  $\beta_{it} = \beta_t = 1$  for all  $i$  (as in equation (6) in Campbell et al. (2001)). We obtain:

$$r_{it} - r_t^{EW} = \varepsilon_{it} - \frac{1}{N_t} \sum_{j=1}^{N_t} \varepsilon_{jt}$$

From the strong law of large numbers, we know that  $\frac{1}{N_t} \sum_{j=1}^{N_t} \varepsilon_{jt} \xrightarrow[N_t \rightarrow \infty]{} 0$  almost surely. In other words, for large  $N_t$ , we have that  $r_{it} - r_t^{EW} \rightarrow \varepsilon_{it}$  and

$$CSV_t^{EW} = \frac{1}{N_t} \sum_{i=1}^{N_t} (r_{it} - r_t^{EW})^2 \xrightarrow[N_t \rightarrow \infty]{} \frac{1}{N_t} \sum_{i=1}^{N_t} \varepsilon_{it}^2 \simeq \sigma_\varepsilon^2(t) \text{ almost surely} \quad (56)$$

In other words, we draw a direct relationship between the dynamics of the cross-sectional dispersion of realized returns and the dynamics of average idiosyncratic variance. Section 2.2 analyzes a situation with a non-trivial dispersion of beta parameters across stocks. Together with the results of section 3.2, suggests that the homogeneous beta assumptions does not represent a material problem for the CSV as an estimator of idiosyncratic variance.

## B Taking into Account the Cross-Sectional Betas dispersion: The general case

The assumption that  $\beta_{it} = \beta_t$  for all  $i$  is obviously a simplistic one and is done only for exposure purposes. Using the single factor decomposition on the definition of the CSV we have,

$$\begin{aligned} CSV_t^{(w_t)} &= \sum_{i=1}^{N_t} w_{it} \left( r_{it} - r_t^{(w_t)} \right)^2 \\ &= \sum_{i=1}^{N_t} w_{it} \left[ \left( \beta_{it} - \sum_{j=1}^{N_t} w_{jt} \beta_{jt} \right) F_t + \varepsilon_{it} - \sum_{j=1}^{N_t} w_{jt} \varepsilon_{jt} \right]^2 \\ &= F_t^2 \sum_{i=1}^{N_t} w_{it} \left( \beta_{it} - \sum_{j=1}^{N_t} w_{jt} \beta_{jt} \right)^2 + \sum_{i=1}^{N_t} w_{it} \left( \varepsilon_{it} - \sum_{j=1}^{N_t} w_{jt} \varepsilon_{jt} \right)^2 + \\ &\quad 2F_t \sum_{i=1}^{N_t} w_{it} \left( \beta_{it} - \sum_{j=1}^{N_t} w_{jt} \beta_{jt} \right) \left( \varepsilon_{it} - \sum_{j=1}^{N_t} w_{jt} \varepsilon_{jt} \right) \end{aligned}$$

After simple rearrangement of terms we get

$$\begin{aligned} CSV_t^{(w_t)} &= F_t^2 \sum_{i=1}^{N_t} w_{it} \left( \beta_{it} - \sum_{j=1}^{N_t} w_{jt} \beta_{jt} \right)^2 + \sum_{i=1}^{N_t} w_{it} \varepsilon_{it}^2 - \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} w_{jt} w_{it} \varepsilon_{it} \varepsilon_{jt} \\ &\quad + 2F_t \sum_{i=1}^{N_t} w_{it} \varepsilon_{it} \left( \beta_{it} - \sum_{j=1}^{N_t} w_{jt} \beta_{jt} \right) \end{aligned}$$

Denote the cross-sectional variance of stock betas as  $CSV_t^\beta = \sum_{i=1}^{N_t} w_{it} \left( \beta_{it} - \sum_{j=1}^{N_t} w_{jt} \beta_{jt} \right)^2$ . Applying the expectation operator and assuming a strict factor model, the last expression simplifies to<sup>33</sup>,

$$E \left[ CSV_t^{(w_t)} \right] = E \left[ F_t^2 CSV_t^\beta \right] + \sum_{i=1}^{N_t} w_{it} \sigma_{\varepsilon_{it}}^2 - \sum_{i=1}^{N_t} w_{it}^2 \sigma_{\varepsilon_{it}}^2$$

under an equal-weighting scheme,

$$E \left[ CSV_t^{EW} \right] = \left( 1 - \frac{1}{N_t} \right) \frac{1}{N_t} \sum_{i=1}^{N_t} \sigma_{\varepsilon_{it}}^2$$

when  $N_t \rightarrow \infty$ ,

$$E \left[ CSV_t^{EW} \right] \xrightarrow{N_t \rightarrow \infty} E \left[ F_t^2 CSV_t^\beta \right] + \frac{1}{N_t} \sum_{i=1}^{N_t} \sigma_{\varepsilon_{it}}^2$$

## C Variance of the CSV estimator with a general factor structure

Apply the variance operator to equation (20),

$$Var \left( CSV_t^{(w_t)} \right) = Var \left( \varepsilon_t' \Lambda_t \varepsilon_t \right) + Var \left( \varepsilon_t' \Omega_t \varepsilon_t \right) - 2Cov \left( \varepsilon_t' \Lambda_t \varepsilon_t; \varepsilon_t' \Omega_t \varepsilon_t \right) \quad (57)$$

Replacing results (22) (23) and (24) in equation (57) yields

$$Var \left( CSV_t^{(w_t)} \right) = 2 \sum_{i=1}^{N_t} (w_{it} \sigma_{\varepsilon_{it}}^2)^2 + 2 \left( \sum_{i=1}^{N_t} \sum_{j \neq i}^{N_t} w_{it} w_{jt} \sigma_{\varepsilon_{ijt}} \right)^2 \quad (58)$$

$$- 4 \sum_{k=1}^{N_t} w_{kt} \sigma_{\varepsilon_{kt}}^2 \sum_{i=1}^{N_t} \sum_{j \neq i}^{N_t} w_{it} w_{jt} \sigma_{\varepsilon_{ijt}} \quad (59)$$

In order to compare the CSV variance under different weighting schemes, we now consider the case with a strict factor model  $\rho_{ijt}^\varepsilon = 0 \forall i, j$  and with  $\sigma_{\varepsilon_{it}}^2 = \sigma_{\varepsilon_t}^2 \forall i$ . All things being equal, this does not represent a loss of generality. In this context the following Corollary holds.

---

<sup>33</sup>See section 2.3 for the detail on the expected value of the third and second terms. The fourth term disappears since  $E[\varepsilon_i] = 0$

**Lemma 3** *The variance of the CSV is proportional to the concentration index implied by the weighting scheme as,*

$$\text{Var} \left( \text{CSV}_t^{(w_t)} \right) = 2\sigma_{\varepsilon_t}^4 \left( \sum_{i=1}^{N_t} w_{it}^2 \right)$$

*Consequently, the equally-weighted CSV is the most efficient estimator among the class of estimators defined by the CSV under a weighting scheme satisfying  $0 < w_{it} < 1$*

**Proof** of Lemma 3

Consider equation (59) with  $\rho_{ijt}^\varepsilon = \rho_t^\varepsilon \forall i, j$  and  $\sigma_{\varepsilon_{it}}^2 = \sigma_{\varepsilon_t}^2 \forall i$

$$\begin{aligned} \text{Var} \left( \text{CSV}_t^{(w_t)} \right) &= 2 \left( \sigma_{\varepsilon_t}^4 \sum_{i=1}^{N_t} w_{it}^2 + \left( \rho_t^\varepsilon \sigma_{\varepsilon_t}^2 \sum_{i=1}^{N_t} \sum_{j \neq i}^{N_t} w_{it} w_{jt} \right)^2 - 2\rho_{\varepsilon_t} \sigma_{\varepsilon_t}^4 \sum_{i=1}^{N_t} \sum_{j \neq i}^{N_t} w_{it} w_{jt} \right) \\ &= 2 \left( \sigma_{\varepsilon_t}^4 \sum_{i=1}^{N_t} w_{it}^2 + \rho_{\varepsilon_t}^2 \sigma_{\varepsilon_t}^4 \left( 1 - \sum_{i=1}^{N_t} w_{it}^2 \right)^2 - 2\rho_t^\varepsilon \sigma_{\varepsilon_t}^4 \left( 1 - \sum_{i=1}^{N_t} w_{it}^2 \right) \right) \end{aligned}$$

Then we obtain,

$$\text{Var} \left( \text{CSV}_t^{(w_t)} \right) = 2\sigma_{\varepsilon_t}^4 \left( \sum_{i=1}^{N_t} w_{it}^2 + \rho_{\varepsilon_t}^2 \left( 1 - \sum_{i=1}^{N_t} w_{it}^2 \right)^2 - 2\rho_t^\varepsilon \left( 1 - \sum_{i=1}^{N_t} w_{it}^2 \right) \right) \quad (60)$$

Considering a strict factor model as described in Assumption 3, then all communalities are captured by the factor(s) and the idiosyncratic components are independent among stocks with  $\rho_{ijt}^\varepsilon = 0 \forall i \neq j$ , yielding,

$$\text{Var} \left( \text{CSV}_t^{(w_t)} \right) = 2\sigma_{\varepsilon_t}^4 \left( \sum_{i=1}^{N_t} w_{it}^2 \right) \quad (61)$$

By definition, a weighting scheme that satisfies  $0 < w_i < 1$ , implies that the term  $\sum_{i=1}^{N_t} w_{it}^2$  is minimum for  $w_{it} = 1/N_t \forall i$ . Using the equal-weighted scheme in equation (61) we have,

$$\text{Var} \left( \text{CSV}_t^{(EW)} \right) = 2\sigma_{\varepsilon_t}^4 \left( \frac{1}{N_t} \right) \quad \blacksquare \quad (62)$$



## D Tables and Figures

Table 1: **Impact of the Betas' Cross-Sectional Dispersion:** This table contains a summary of the distribution of the following time series: the cross-sectional dispersion of betas  $CSV_t^\beta$ , estimated with respect to the CAPM at the end of the every month using daily time series; the product of the square return of the market portfolio  $F_t^2$  and  $CSV_t^\beta$ ; the average idiosyncratic variance  $\sigma_{\varepsilon_t}^2$  and the proportion of the product  $F_t^2 CSV_t^\beta$  to  $\sigma_{\varepsilon_t}^2$ . This illustrates the magnitude of the bias introduced by the assumption of null cross-sectional dispersion of betas. The upper part of the table contains the beta estimates based on an equally-weighted market portfolio and an equal-weighted average Idiosyncratic variance. The lower part of the table presents the corresponding series for a market capitalization weighting scheme. The  $\sigma_{\varepsilon_t}^2$  is presented in annualized terms. The period is July 1963 to December 2006.

| Equal-Weighted                                       | $Q_{2.5}$ | $Q_{25}$ | $Q_{50}$ | $Q_{75}$ | $Q_{97.5}$ |
|--|-----------|----------|----------|----------|------------|
| $CSV_t^\beta$  | 0.72      | 1.38     | 2.23     | 4.11     | 11.17      |
| $F_t^2 CSV_t^\beta$                                  | 0.00%     | 0.07%    | 0.33%    | 1.05%    | 3.88%      |
| $\sigma_{\varepsilon_t}^2$                           | 0.11      | 0.16     | 0.24     | 0.59     | 1.02       |
| $\frac{F_t^2 CSV_t^\beta}{\sigma_{\varepsilon_t}^2}$ | 0.00%     | 0.15%    | 0.65%    | 2.08%    | 6.67%      |
| Cap-Weighted   |           |          |          |          |            |
| $CSV_t^\beta$  | 0.22      | 0.37     | 0.48     | 0.67     | 1.27       |
| $F_t^2 CSV_t^\beta$                                  | 0.00%     | 0.01%    | 0.04%    | 0.12%    | 0.39%      |
| $\sigma_{\varepsilon_t}^2$                           | 0.02      | 0.05     | 0.07     | 0.09     | 0.20       |
| $\frac{F_t^2 CSV_t^\beta}{\sigma_{\varepsilon_t}^2}$ | 0.00%     | 0.09%    | 0.31%    | 0.97%    | 2.81%      |

Table 2: **Homogeneous Betas Bias estimation:** This table contains the output summary of the regression  $CSV^{wt} = bias + \psi\sigma_{model}^2(w_t) + \zeta$ . Where  $\sigma_{model}^2(w_t)$  represents monthly estimates of the weighted average idiosyncratic variance estimated using the corresponding model (either CAPM or FF). The weighting scheme for the average is taken consistent with the weighting scheme of the  $CSV$  and is either cap-weighted (CW) or equal-weighted (EW). The period is July 1963 to December 2006.

| <b>Betas bias</b> | $CAPM^{EW}$ | $FF^{EW}$ | $CAPM^{CW}$ | $FF^{CW}$ |
|-------------------|-------------|-----------|-------------|-----------|
| bias              | 1.1e-005    | 2.2e-005  | -2.0e-005   | -3.7e-005 |
| NW t-stat         | 1.757       | 2.369     | -2.731      | -4.708    |
| Std               | 3.0e-006    | 5.9e-006  | 2.0e-006    | 3.5e-006  |
| $\psi$            | 0.983       | 0.988     | 1.124       | 1.243     |
| NW t-stat         | 154.808     | 100.062   | 39.161      | 38.802    |
| Std               | 0.002       | 0.003     | 0.005       | 0.009     |
| $\bar{R}^2$ (%)   | 99.868      | 99.494    | 98.990      | 97.064    |

Table 3: **Summary Statistics Monthly Idiosyncratic Variance:** This table contains the mean and standard deviation of the monthly time series for: the CSV, the average idiosyncratic variance based on the CAPM and the FF models and the average (mostly idiosyncratic) variance measure ala Goyal and Santa-Clara (2003) as in equations (4), (28) and (29) using both, equal-weighted and cap-weighted schemes. The period is January 1964 to December 2006.

|          | $CSV^{EW}$ | $FF^{EW}$ | $CAPM^{EW}$ | $GS^{EW}$ |
|----------|------------|-----------|-------------|-----------|
| Mean     | 38.42%     | 38.32%    | 38.79%      | 34.24%    |
| Std.Dev. | 8.52%      | 8.60%     | 8.66%       | 6.99%     |
|          | $CSV^{CW}$ | $FF^{CW}$ | $CAPM^{CW}$ | $GS^{CW}$ |
| Mean     | 8.52%      | 7.62%     | 8.04%       | 11.23%    |
| Std.Dev. | 2.03%      | 1.61%     | 1.80%       | 2.91%     |

Table 4: **Correlations Monthly Time Series:** This table contains the cross-correlation for the monthly time series of: the CSV, the average idiosyncratic variance based on the CAPM and the FF models and the average (mostly idiosyncratic) variance measure ala Goyal and Santa-Clara (2003) as in equations (4), (28) and (29) using both, equal-weighted and cap-weighted schemes. The period is January 1964 to December 2006.

|   | $CSV^{EW}$ | $FF^{EW}$ | $CAPM^{EW}$ | $GS^{EW}$ | $CSV^{CW}$ | $FF^{CW}$ | $CAPM^{CW}$ | $GS^{CW}$ |
|---|------------|-----------|-------------|-----------|------------|-----------|-------------|-----------|
| 1 | 99.75%     | 99.93%    | 95.47%      | 72.33%    | 75.20%     | 73.08%    | 61.49%      |           |
|   | 1          | 99.86%    | 93.71%      | 68.68%    | 72.38%     | 69.70%    | 56.98%      |           |
|   |            | 1         | 94.89%      | 70.89%    | 73.95%     | 71.81%    | 60.27%      |           |
|   |            |           | 1           | 82.47%    | 83.04%     | 82.54%    | 76.75%      |           |
|   |            |           |             | 1         | 98.52%     | 99.49%    | 92.67%      |           |
|   |            |           |             |           | 1          | 99.19%    | 88.19%      |           |
|   |            |           |             |           |            | 1         | 92.04%      |           |

Table 5: **Summary Statistics Daily Time Series:** This table contains the mean and standard deviation of the daily time series for: the CSV and the average idiosyncratic variance based on the FF model as in equations (4) and (28) using both, equal-weighted and cap-weighted schemes. The period is January 1964 to December 2006.

| Statistic | $CSV^{EW}$ | $FF^{EW}$ | $CSV^{CW}$ | $FF^{CW}$ |
|-----------|------------|-----------|------------|-----------|
| Mean      | 38.44%     | 38.22%    | 8.52%      | 7.77%     |
| Std.Dev.  | 9.79%      | 8.54%     | 2.48%      | 1.65%     |

Table 6: **Correlations Daily Time Series:** This table contains the cross-correlation of the daily time series for: the CSV and the average idiosyncratic variance based on the FF model as in equations (4) and (28) using both, equal-weighted and cap-weighted schemes. The period is January 1964 to December 2006.

|            | $CSV^{EW}$ | $FF^{EW}$ | $CSV^{CW}$ | $FF^{CW}$ |
|------------|------------|-----------|------------|-----------|
| $CSV^{EW}$ | 1          | 81.41%    | 61.17%     | 62.31%    |
| $FF^{EW}$  |            | 1         | 52.98%     | 72.55%    |
| $CSV^{CW}$ |            |           | 1          | 73.67%    |

Table 7: **Regime Switching Parameters Monthly Series:** This table contains the parameter estimates of the Markov regime-switching model specified in equation 31 using the monthly time series of the CSV and the average idiosyncratic variance based on the FF model as in equations (4) and (28) using both, equal-weighted and cap-weighted schemes.  $\mu_i$  is the average level of the variable on regime  $i$ ,  $\sigma_i$  is the standard deviation level of the variable on regime  $i$ ,  $\phi$  is the autocorrelation coefficient, p and q are the probabilities of remaining in regimes 1 and 2 correspondingly. The period is January 1964 to December 2006.

|            | $FF^{CW}$ | $CSV^{CW}$ | $FF^{EW}$ | $CSV^{EW}$ |
|------------|-----------|------------|-----------|------------|
| $\mu_1$    | 11.48%    | 10.69%     | 36.04%    | 39.91%     |
| $\mu_2$    | 6.09%     | 6.54%      | 27.33%    | 29.71%     |
| $\sigma_1$ | 7.33%     | 9.99%      | 21.40%    | 23.29%     |
| $\sigma_2$ | 1.13%     | 1.25%      | 3.09%     | 3.46%      |
| $\phi$     | 83.92%    | 83.86%     | 98.13%    | 98.05%     |
| p          | 90.64%    | 85.71%     | 82.24%    | 83.87%     |
| q          | 98.99%    | 98.00%     | 95.10%    | 96.26%     |

Table 8: **Regime Switching Parameters Daily Series:** This table contains the parameter estimates of the Markov regime-switching model specified in equation 31 using the daily time series of the CSV and the average idiosyncratic variance based on the FF model as in equations (4) and (28) using both, equal-weighted and cap-weighted schemes.  $\mu_i$  is the average level of the variable on regime  $i$ ,  $\sigma_i$  is the standard deviation level of the variable on regime  $i$ ,  $\phi$  is the autocorrelation coefficient, p and q are the probabilities of remaining in regimes 1 and 2 correspondingly. The period is January 1964 to December 2006.

|            | $FF^{CW}$ | $CSV^{CW}$ | $FF^{EW}$ | $CSV^{EW}$ |
|------------|-----------|------------|-----------|------------|
| $\mu_1$    | 2.69%     | 10.96%     | 12.91%    | 44.56%     |
| $\mu_2$    | 2.68%     | 6.41%      | 12.92%    | 30.40%     |
| $\sigma_1$ | 0.70%     | 14.56%     | 2.97%     | 62.06%     |
| $\sigma_2$ | 0.10%     | 1.49%      | 0.32%     | 4.79%      |
| $\phi$     | 99.99%    | 82.54%     | 99.98%    | 96.40%     |
| p          | 89.23%    | 77.83%     | 83.96%    | 71.05%     |
| q          | 98.00%    | 97.04%     | 96.12%    | 95.41%     |

Table 9: **Cross-sectional Skewness Distribution:** This table contains the minimum, maximum and the 2.5%, 25%, 50%, 75% and 97.5% quantiles of the distribution for the time series of skewness using conventional (EW), cap-weighted (CW) and robust estimates as in equations (4) and (42). The monthly series are calculated as the average over the month of the daily estimates. The period is July 1963 to December 2006.

| <b>Daily Frequency</b> | Min     | $Q_{2.5}$ | $Q_{25}$ | $Q_{50}$ | $Q_{75}$ | $Q_{97.5}$ | Max     |
|------------------------|---------|-----------|----------|----------|----------|------------|---------|
| $CSS^{EW}$             | -5.369  | -0.917    | 0.592    | 1.387    | 2.462    | 10.149     | 68.239  |
| $CSS^{CW}$             | -20.252 | -2.119    | -0.155   | 0.569    | 1.180    | 3.832      | 176.908 |
| Robust                 | -1.000  | -1.000    | -0.220   | 0.026    | 0.252    | 1.000      | 1.000   |

| <b>Monthly Frequency</b> | Min    | $Q_{2.5}$ | $Q_{25}$ | $Q_{50}$ | $Q_{75}$ | $Q_{97.5}$ | Max    |
|--------------------------|--------|-----------|----------|----------|----------|------------|--------|
| $CSS^{EW}$               | -0.339 | 0.173     | 1.066    | 1.719    | 2.702    | 6.482      | 9.115  |
| $CSS^{CW}$               | -1.376 | -0.533    | 0.159    | 0.592    | 0.960    | 2.760      | 10.999 |
| Robust                   | -0.449 | -0.304    | -0.055   | 0.017    | 0.092    | 0.342      | 0.533  |

Table 10: **Predictability Regression on CW market portfolio using EW average measures of idiosyncratic risk:** This table presents the results of a one-month ahead predictive regression of the excess cap-weighted monthly portfolio returns, denoted by  $r^{CW}$ , on the monthly lagged equal-weighted average idiosyncratic variance for two different measures and for three sample periods.  $CAPM^{EW}$  is the average idiosyncratic variance derived from CAPM and  $CSV^{EW}$  is the cross-sectional variance estimated as the average of the daily estimations of the month obtained as in equation (4). The intercept, the regression coefficient of the corresponding lagged idiosyncratic variance denoted by  $\beta$ , the standard error of the regression coefficients denoted by Std, the Newey-West autocorrelation corrected t-stats and the adjusted coefficient of determination denoted by  $\bar{R}^2$  are reported. The sample periods are 1963:08 - 1999:12, 1963:08 - 2001:12 and 1963:08 - 2006:12.

| <b>Monthly series</b>                  | 1963:08 - 1999:12 |            | 1963:08 - 2001:12 |            | 1963:08 - 2006:12 |            |
|--|-------------------|------------|-------------------|------------|-------------------|------------|
| <b>Forecasting <math>r^{CW}</math></b> | $CAPM^{EW}$       | $CSV^{EW}$ | $CAPM^{EW}$       | $CSV^{EW}$ | $CAPM^{EW}$       | $CSV^{EW}$ |
| Intercept                              | -0.001            | -0.002     | 8.1e-004          | 9.4e-004   | 0.002             | 0.002      |
| NW t-stat                              | -0.442            | -0.504     | 0.228             | 0.257      | 0.676             | 0.696      |
| Std                                    | 0.003             | 0.003      | 0.003             | 0.003      | 0.003             | 0.003      |
| <i>Coefficient</i>                     | 0.291             | 0.302      | 0.169             | 0.167      | 0.129             | 0.126      |
| NW t-stat                              | 4.255             | 4.334      | 1.744             | 1.623      | 1.398             | 1.305      |
| Std                                    | 0.090             | 0.093      | 0.081             | 0.082      | 0.077             | 0.078      |
| $\bar{R}^2(\%)$                        | 2.106             | 2.148      | 0.728             | 0.673      | 0.349             | 0.312      |

Table 11: **Predictability Regression on EW market portfolio with EW average measures of idiosyncratic risk:** This table presents the results of a one-month ahead predictive regression of the excess equal-weighted monthly portfolio returns, denoted by  $r^{EW}$ , on the monthly lagged equal-weighted average idiosyncratic variance for two different measures and for three sample periods.  $CAPM^{EW}$  is the average idiosyncratic variance derived from CAPM and  $CSV^{EW}$  is the cross-sectional variance estimated as the average of the daily estimations of the month obtained as in equation (4). The intercept, the regression coefficient of the corresponding lagged idiosyncratic variance denoted by  $\beta$ , the standard error of the regression coefficients denoted by Std, the Newey-West autocorrelation corrected t-stats and the adjusted coefficient of determination denoted by  $\bar{R}^2$  are reported. The sample periods are 1963:08 - 1999:12, 1963:08 - 2001:12 and 1963:08 - 2006:12.

| <b>Monthly series</b>                  | 1963:08 - 1999:12 |            | 1963:08 - 2001:12 |            | 1963:08 - 2006:12 |            |
|--|-------------------|------------|-------------------|------------|-------------------|------------|
| <b>Forecasting <math>r^{EW}</math></b> | $CAPM^{EW}$       | $CSV^{EW}$ | $CAPM^{EW}$       | $CSV^{EW}$ | $CAPM^{EW}$       | $CSV^{EW}$ |
| Intercept                              | 0.002             | 0.002      | 0.001             | 0.001      | 0.002             | 0.003      |
| NW t-stat                              | 0.370             | 0.336      | 0.208             | 0.238      | 0.543             | 0.569      |
| Std                                    | 0.004             | 0.004      | 0.004             | 0.004      | 0.004             | 0.004      |
| <i>Coefficient</i>                     | 0.265             | 0.274      | 0.260             | 0.259      | 0.236             | 0.235      |
| NW t-stat                              | 2.317             | 2.341      | 2.603             | 2.537      | 2.484             | 2.428      |
| Std                                    | 0.118             | 0.121      | 0.107             | 0.109      | 0.102             | 0.104      |
| $\bar{R}^2(\%)$                        | 0.917             | 0.927      | 1.060             | 1.002      | 0.833             | 0.789      |

Table 12: **Predictability Regression on CW market portfolio with CW idiosyncratic variance measures:** This table presents the results of a one-month ahead predictive regression of the excess cap-weighted monthly portfolio returns, denoted by  $r^{CW}$ , on the monthly lagged cap-weighted idiosyncratic variance for two different measures and for three sample periods.  $CAPM^{CW}$  is the cap-weighted idiosyncratic variance derived from CAPM and  $CSV^{CW}$  is the cross-sectional variance estimated as the average of the daily estimations of the month obtained as in equation (5). The intercept, the regression coefficient of the corresponding lagged idiosyncratic variance denoted by  $\beta$ , the standard error of the regression coefficients denoted by Std, the Newey-West autocorrelation corrected t-stats and the adjusted coefficient of determination denoted by  $\bar{R}^2$  are reported. The sample periods are 1963:08 - 1999:12, 1963:08 - 2001:12 and 1963:08 - 2006:12.

| <b>Monthly series</b>                  | 1963:08 - 1999:12 |            | 1963:08 - 2001:12 |            | 1963:08 - 2006:12 |            |
|--|-------------------|------------|-------------------|------------|-------------------|------------|
| <b>Forecasting <math>r^{CW}</math></b> | $CAPM^{CW}$       | $CSV^{CW}$ | $CAPM^{CW}$       | $CSV^{CW}$ | $CAPM^{CW}$       | $CSV^{CW}$ |
| Intercept                              | 6.7e-004          | 0.002      | 0.007             | 0.008      | 0.008             | 0.009      |
| NW t-stat                              | 0.151             | 0.370      | 2.200             | 2.449      | 2.702             | 2.966      |
| Std                                    | 0.004             | 0.004      | 0.003             | 0.003      | 0.003             | 0.003      |
| <i>Coefficient</i>                     | 1.098             | 0.890      | -0.169            | -0.217     | -0.274            | -0.309     |
| NW t-stat                              | 1.478             | 1.232      | -0.370            | -0.550     | -0.663            | -0.865     |
| Std                                    | 0.680             | 0.625      | 0.387             | 0.341      | 0.370             | 0.328      |
| $\bar{R}^2$ (%)                        | 0.368             | 0.236      | -0.176            | -0.130     | -0.087            | -0.021     |

Table 13: **Daily predictability Regression on EW portfolio with EW average idiosyncratic variance measures:** This table presents the results of a one-day ahead predictive regression of the excess equal-weighted daily portfolio returns, denoted by  $r^{EW}$ , on the daily lagged equal-weighted cross-sectional variance denoted as  $CSV^{EW}$  estimated as in equation (4) for three sample periods. The intercept, the regression coefficient corresponding to the CSV, the standard error of the regression coefficients denoted by std, the Newey-West autocorrelation corrected (30 lags) t-stats and the adjusted coefficient of determination denoted by  $\bar{R}^2$  are reported. The sample periods are 1963:07 to 1999:12, 1963:07 to 2001:12 and 1963:07 to 2006:12.

| Daily series                           | 63:07-99:12 | 63:07-01:12 | 63:07-06:12 |
|--|-------------|-------------|-------------|
| <b>Forecasting <math>r^{EW}</math></b> | $CSV^{EW}$  | $CSV^{EW}$  | $CSV^{EW}$  |
| Intercept                              | -0.000      | -0.000      | -0.000      |
| NW t-stat                              | -0.896      | -0.816      | -0.102      |
| Std                                    | 0.000       | 0.000       | 0.000       |
| $\beta$                                | 0.582       | 0.521       | 0.442       |
| NW t-stat                              | 4.943       | 4.666       | 4.044       |
| Std                                    | 0.062       | 0.057       | 0.054       |
| $\bar{R}^2$ (%)                        | 0.932       | 0.835       | 0.591       |



Table 14: **Daily predictability Regression on CW portfolio with CW average idiosyncratic variance measures:** This table presents the results of a one-day ahead predictive regression of the excess equal-weighted daily portfolio returns, denoted by  $r^{CW}$ , on the daily lagged equal-weighted cross-sectional variance denoted as  $CSV^{CW}$  estimated as in equation (4) for three sample periods. The intercept, the regression coefficient corresponding to the CSV, the standard error of the regression coefficients denoted by std, the Newey-West autocorrelation corrected (30 lags) t-stats and the adjusted coefficient of determination denoted by  $\bar{R}^2$  are reported. The sample periods are 1963:07 to 1999:12, 1963:07 to 2001:12 and 1963:07 to 2006:12.

| <b>Daily series</b>                    | 63:07-99:12 | 63:07-01:12 | 63:07-06:12 |
|--|-------------|-------------|-------------|
| <b>Forecasting <math>r^{CW}</math></b> | $CSV^{CW}$  | $CSV^{CW}$  | $CSV^{CW}$  |
| Intercept                              | -0.001      | -0.000      | -0.000      |
| NW t-stat                              | -3.454      | -0.824      | -0.809      |
| Std                                    | 0.000       | 0.000       | 0.000       |
| $\beta$                                | 3.775       | 1.389       | 1.358       |
| NW t-stat                              | 5.724       | 2.039       | 2.083       |
| Std                                    | 0.396       | 0.259       | 0.256       |
| $\bar{R}^2$ (%)                        | 0.966       | 0.286       | 0.248       |

Table 15: **Daily and Monthly predictability with and without market volatility:** This table presents summary statistics for three regressions. The upper panel corresponds to daily estimations and the lower panel its counterpart on the monthly predictions. In each panel, the first row corresponds to the regression  $r_t^{EW} = \alpha + \beta CSV_t + \epsilon_t$  the second row to  $r_t^{EW} = \alpha + \vartheta Var(r_t^{EW}) + \epsilon_t$  and the third one to  $r_t^{EW} = \alpha + \beta CSV_t + \vartheta Var(r_t^{EW}) + \epsilon_t$ . The sample period is July 1963 to December 2006

| <b>Daily Estimates</b>   | Intercept | NW t-stat | CSV   | NW t-stat | $Var(r^{EW})$ | NW t-stat | $\overline{R}^2(\%)$ |
|--------------------------|-----------|-----------|-------|-----------|---------------|-----------|----------------------|
| Forecasting $r^{EW}$     | -0.000    | -0.102    | 0.442 | 4.044     |               |           | 0.591                |
| Forecasting $r^{EW}$     | 0.001     | 4.928     |       |           | -1.496        | -0.656    | 0.010                |
| Forecasting $r^{EW}$     | 0.000     | 0.518     | 0.502 | 4.647     | -4.173        | -2.115    | 0.717                |
| <b>Monthly Estimates</b> | Intercept | NW t-stat | CSV   | NW t-stat | $Var(r^{EW})$ | NW t-stat | $\overline{R}^2(\%)$ |
| Forecasting $r^{EW}$     | 0.003     | 0.569     | 0.235 | 2.428     |               |           | 0.789                |
| Forecasting $r^{EW}$     | 0.010     | 3.276     |       |           | 0.073         | 0.053     | -0.192               |
| Forecasting $r^{EW}$     | 0.003     | 0.628     | 0.253 | 2.456     | -0.742        | -0.642    | 0.668                |

Table 16: **Predictability Regression on EW portfolio with right and left  $CSV^{EW}$  average idiosyncratic variance measures:** This table presents the results of a one-day and one-month ahead predictive regression of the excess equal-weighted portfolio returns, denoted by  $r^{EW}$ , on the daily or monthly (correspondingly) lagged equal-weighted cross-sectional variance of the returns to the right (higher than) of the cross-sectional distribution mean (which is actually  $r_t^{EW}$ ) denoted as  $CSV^+$  and the cross-sectional variance of the returns to the left (lower than) the mean of the cross-sectional distribution  $r_t^{EW}$ , denoted as  $CSV^-$ . The intercept, the regression coefficients corresponding to the  $CSV^+$  and  $CSV^-$ , the standard error of the regression coefficients denoted by Std, the Newey-West autocorrelation corrected t-stats and the adjusted coefficient of determination denoted by  $\bar{R}^2$  are reported. The sample period is 1963:07 to 2006:12.

| <b>Forecasting</b> | $r^{EW}$ | $Monthly^{EW}$ | $Daily^{EW}$ |
|--------------------|----------|----------------|--------------|
| Intercept          |          | 0.005          | 6.8e-004     |
| NW t-stat          |          | 0.959          | 3.828        |
| Std                |          | 0.004          | 1.2e-004     |
| $CSV^+$            |          | 0.377          | 0.490        |
| NW t-stat          |          | 1.999          | 3.292        |
| Std                |          | 0.166          | 0.040        |
| $CSV^-$            |          | -0.440         | -1.108       |
| NW t-stat          |          | -0.783         | -3.174       |
| Std                |          | 0.444          | 0.151        |
| $\bar{R}^2$ (%)    |          | 1.178          | 1.368        |

Table 17: **Daily predictability with skewness for  $r^{EW}$** : This table presents the results of one-day ahead predictive regressions of the excess equal-weighted daily portfolio returns, denoted by  $r^{EW}$ . The first explanatory variable is the daily lagged estimate of the equal-weighted CSV estimated as in equation (4); The second explanatory variable is the robust estimate of skewness estimated as in equations (42). The intercept, the corresponding regression coefficients together with their Newey-West autocorrelation corrected t-stats with 30 lags and standard errors are reported.  $\bar{R}^2$  denotes adjusted coefficient of determination. The regression is reported for three different sample periods: 1963:07 to 1999:12, 1963:07 to 2001:12, 1963:07 to 2006:12.

| <b>1963:07-2006:12</b> | Coeff.    | t-stat | Std.Dev. | $\bar{R}^2$ (%) |
|------------------------|-----------|--------|----------|-----------------|
| Intercept              | -3.7e-005 | -0.234 | 0.000    | 5.833           |
| $CSV^{EW}$             | 0.402     | 4.013  | 0.053    |                 |
| Skewness               | 0.004     | 20.190 | 0.000    |                 |
| <b>1963:07-2001:12</b> |           |        |          |                 |
| Intercept              | -1.7e-004 | -1.056 | 1.1e-004 | 6.800           |
| $CSV^{EW}$             | 0.477     | 4.717  | 0.056    |                 |
| Skewness               | 0.004     | 20.040 | 1.5e-004 |                 |
| <b>1963:07-1999:12</b> |           |        |          |                 |
| Intercept              | -1.8e-004 | -1.095 | 1.1e-004 | 7.442           |
| $CSV^{EW}$             | 0.522     | 5.008  | 0.060    |                 |
| Skewness               | 0.004     | 19.864 | 1.4e-004 |                 |

Table 18: **Monthly predictability with skewness for  $r^{EW}$** : This table presents the results of one-month ahead predictive regressions of the excess equal-weighted monthly portfolio returns, denoted by  $r^{EW}$ . The first explanatory variable is the monthly lagged equal-weighted CSV and the second explanatory variable is the lagged robust skewness, both estimated as the average at the end of the month of their corresponding daily estimates as in equations (4) and (42). The intercept, corresponding regression coefficients together with Newey-West autocorrelation corrected t-stats with 12 lags and standard errors are reported.  $\bar{R}^2$  denotes adjusted coefficient of determination. The regression is reported for three different sample periods: 1963:07 to 1999:12, 1963:07 to 2001:12, 1963:07 to 2006:12.

| <b>1963:07-2006:12</b> | Coeff. | t-stat | Std.Dev. | $\bar{R}^2$ (%) |
|------------------------|--------|--------|----------|-----------------|
| Intercept              | 0.000  | 0.107  | 0.004    | 4.587           |
| $CSV^{EW}$             | 0.250  | 2.518  | 0.102    |                 |
| Skewness               | 0.078  | 4.458  | 0.017    |                 |

| <b>1963:07-2001:12</b> |           |        |       |       |
|------------------------|-----------|--------|-------|-------|
| Intercept              | -9.2e-004 | -0.205 | 0.004 | 4.847 |
| $CSV^{EW}$             | 0.271     | 2.644  | 0.107 |       |
| Skewness               | 0.075     | 4.287  | 0.017 |       |

| <b>1963:07-1999:12</b> |           |        |       |       |
|------------------------|-----------|--------|-------|-------|
| Intercept              | -8.8e-005 | -0.020 | 0.004 | 5.076 |
| $CSV^{EW}$             | 0.272     | 2.456  | 0.119 |       |
| Skewness               | 0.073     | 4.123  | 0.016 |       |

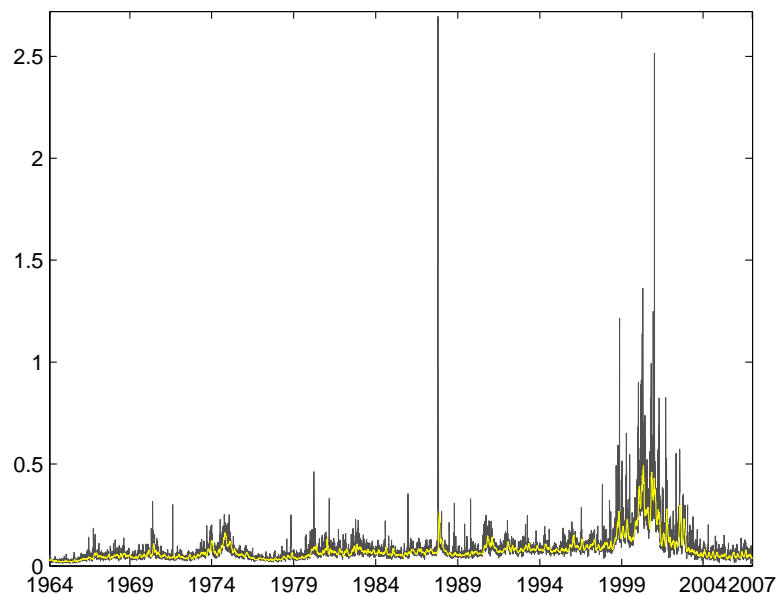


Figure 1: Annualized cap-weighted idiosyncratic variances, daily estimation: The yellow line is the time series of the cap-weighted idiosyncratic variance with respect to the FF model estimated daily as in equation 28. The darker line shows the time series of the cap-weighted version of CSV estimated daily as in equation 5. The sample period is January 1964 up to December 2006.

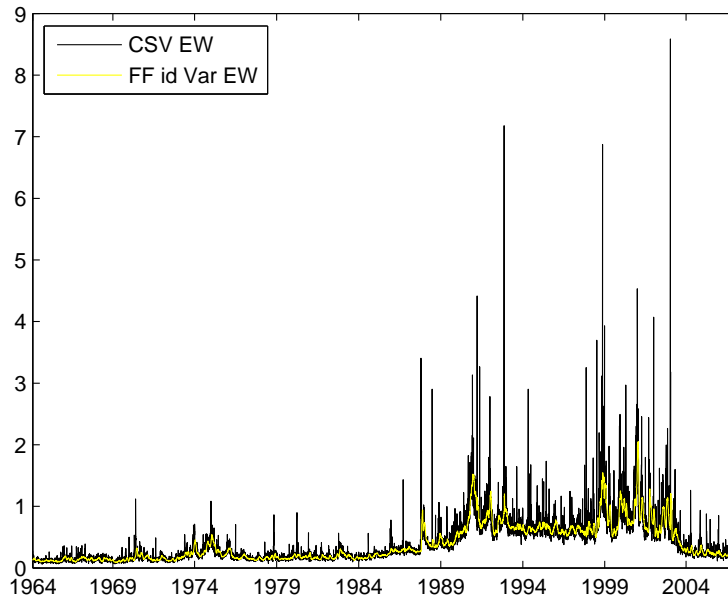


Figure 2: Annualized equally-weighted idiosyncratic variances, daily estimation: The yellow line is the time series of the equal-weighted average idiosyncratic variance with respect to the FF model estimated daily similar to equation 28. The darker line shows the time series of the  $CSV^{EW}$  estimated daily as in equation 4. The sample period is January 1964 up to December 2006.

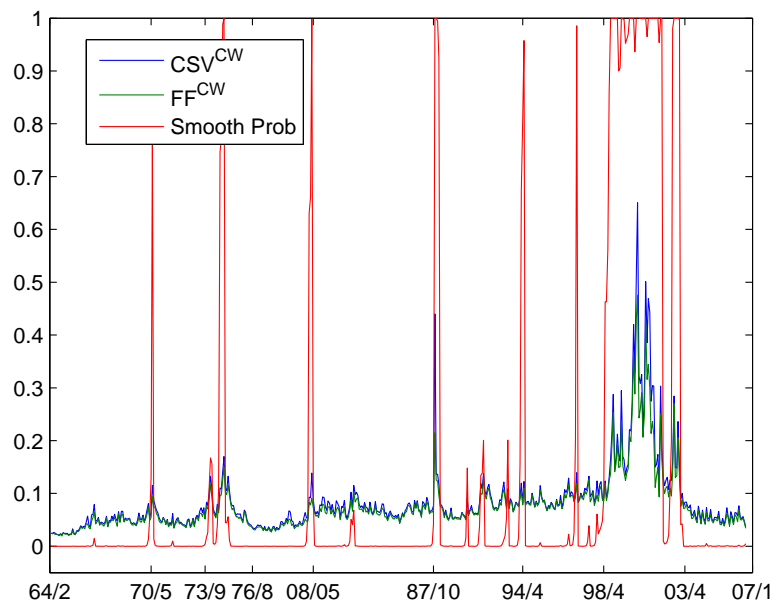


Figure 3: Smoothed probability and annualized cap-weighted  $CSV$  and FF cap-weighted Idiosyncratic Variance, monthly estimation: The red line plots the Smoothed (conditional) probability of the  $CSV^{CW}$  being in the high-mean high-variance regime of a Markov regime-switching model specified in equation 31. The green line plots the cap-weighted idiosyncratic variance with respect to the FF model estimated every month as in equation 28. The blue line shows the monthly time series of the  $CSV^{CW}$  estimated at the end of each month as the average of the daily estimations (as in equation 5) during the month. The sample period is January 1964 up to December 2006. In the x axis appear selected months that mark obvious changes in regime by the smoothed probability.



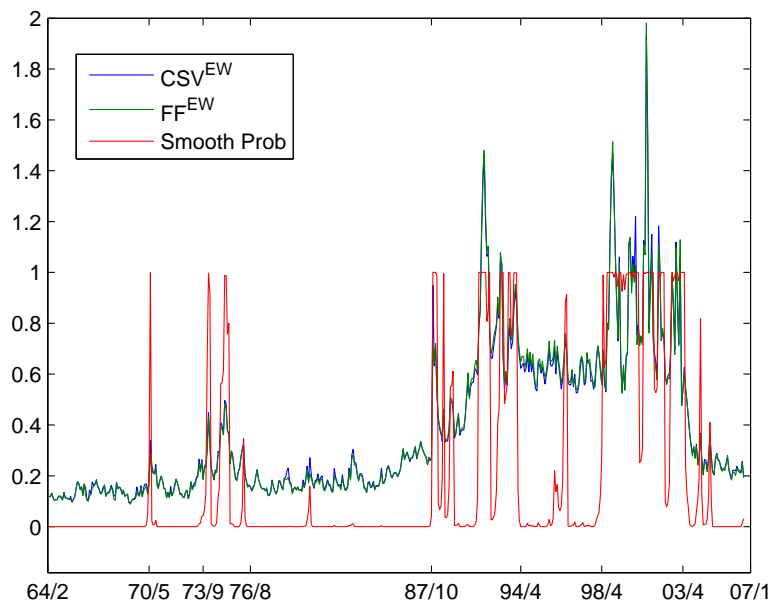


Figure 4: Smoothed probabilities and annualized equal-weighted  $CSV$  and  $FF$  average idiosyncratic variance monthly estimation: The red line plots the Smoothed (conditional) probability of the  $CSV^{EW}$  being in the high-mean high-variance regime of a Markov regime-switching model specified in equation 31. The green line plots the equal-weighted average idiosyncratic variance with respect to the  $FF$  model estimated daily similar to equation 28. The blue line shows the monthly time series of the  $CSV^{EW}$  estimated at the end of each month as the average of the daily estimations (as in equation 4) during the month. The sample period is January 1964 up to December 2006. In the x axis appear selected months that mark obvious changes in regime by the smoothed probability.

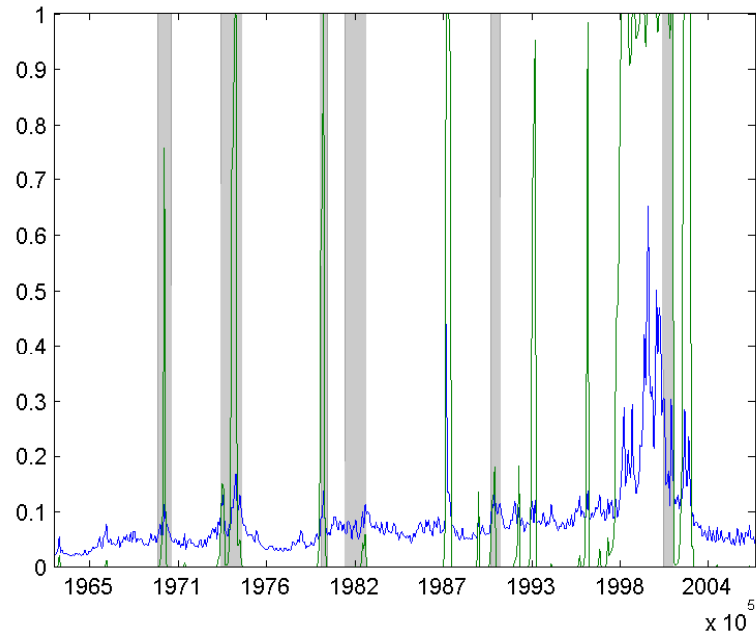


Figure 5: Smoothed probabilities and annualized cap-weighted  $CSV$  monthly estimation: The green line plots the Smoothed (conditional) probability of the  $CSV^{CW}$  being in the high-mean high-variance regime of a Markov regime-switching model specified in equation 31. The blue line shows the monthly time series of the  $CSV^{CW}$  estimated at the end of each month as the average of the daily estimations (as in equation 4) during the month. The shaded areas are the NBER recessions. The sample period is July 1963 to December 2006.

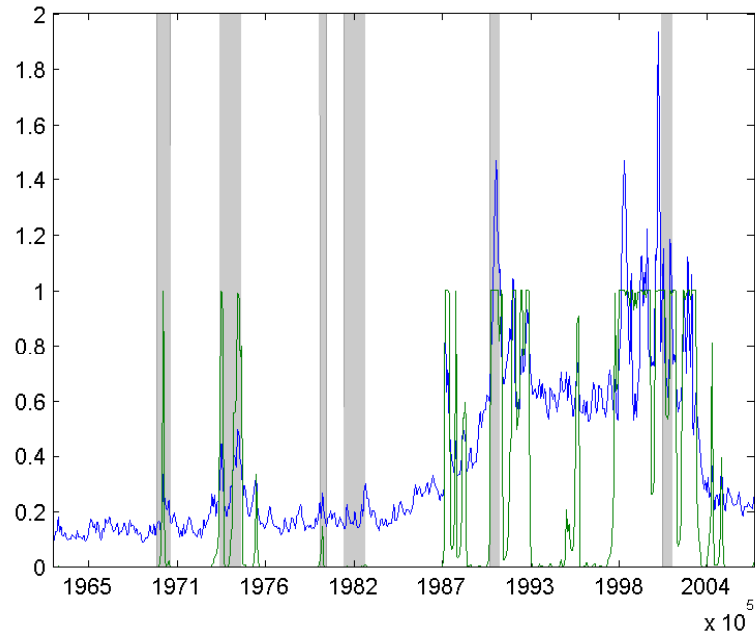


Figure 6: Smoothed probabilities and annualized equal-weighted  $CSV^{EW}$  monthly estimation: The green line plots the Smoothed (conditional) probability of the  $CSV^{EW}$  being in the high-mean high-variance regime of a Markov regime-switching model specified in equation 31. The blue line shows the monthly time series of the  $CSV^{EW}$  estimated at the end of each month as the average of the daily estimations (as in equation 4) during the month. The shaded areas are the NBER recessions. The sample period is July 1963 to December 2006.